Attorney Docket No.: 15303-000310
Client Reference No.: 51836AUST1

PATENT APPLICATION

ESTROGEN-REGULATED UNCONVENTIONAL MYOSIN-RELATED PROTEIN: COMPOSITIONS AND METHODS OF USE

Inventors:

Alan R. Brooks, a citizen of the United Kingdom, residing at,
8018 Kelok Way, Clayton, CA 94517

Gary G. Deng, a citizen of the United States, residing at,
453 Aspinwall Dr., Benicia, CA 94510

Gabor M. Rubanyi, a citizen of the United States, residing at,
2426 Heritage Oaks Dr., Alamo, CA 94507

Assignee:

Schering Aktiengesellschaft

Entity:

Large

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
Tel: 415-576-0200

ESTROGEN-REGULATED UNCONVENTIONAL MYOSIN-RELATED PROTEIN: COMPOSITIONS AND METHODS OF USE

CROSS-REFERENCE TO RELATED APPLICATION

This application claims benefit of U.S. Provisional Application No. 60/188,488, filed March 20, 2000, which application is incorporated herein by reference for all purposes.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

Not applicable.

BACKGROUND OF THE INVENTION

The ovarian sex steroid hormone estrogen has a broad spectrum of biological functions including inhibiting the development of atherosclerosis, osteoporosis, and neurological disorders including Alzheimer's Disease and Parkinson's Disease. Estrogen has also been associated with breast and other cancers. Despite the importance of estrogen, and estrogen signaling, in these and other diseases, the molecular mechanisms of estrogen action are largely unknown, and the ability to modulate estrogen signaling remains poorly developed.

Estrogen regulates the expression of many genes by binding to its two known nuclear receptors, ERα and ERβ. These two receptor types are encoded by distinct genes, and have distinct ligand affinity and tissue distribution (*see, e.g.*, Kuiper *et al.*, *Endocrinology* 1997 138:863-70). The importance of estrogen receptor α has been demonstrated by the generation of a knock out mouse, called ERKO. Both male and female ERKO mice are sterile and display a variety of phenotypic effects including decreased bone density, defects in their reproductive tissues, and decreased likelihood of oncogene-induced cancer. *See, e.g.*, Couse *et al.*, (1999) *Endocr Rev* 20:358-417; Korach (1994) *Science* 266:1524-7; Bocchinfuso *et al.*, (1999) *Cancer Res* 59:1869-76. Mice lacking estrogen receptor β are fertile, but females display decreased ovarian activity, leading to decreased litter sizes. Krege, *et al.* (1998) *Proc Natl Acad Sci U S A* 95:15677-82. Mice lacking both α and β receptors are infertile and display an ovarian phenotype that is distinct from that of either receptor knockout alone. Couse *et al.*, (1999) *Science* 286:2328-31.

Myosins represent a large class of homologous proteins that have been categorized into at least 15 distinct classes. Class II represents conventional myosins, which are two-headed, filament forming proteins that provide the basis for muscle contraction. The remaining classes represent proteins with distinct structures and functions. For example, members of the myosin superfamily are single-headed or two-headed, and contain any number of a variety of structural domains including ATP-binding domains, actin-binding domains, SH3 domains, talin domains, MyTH4 domains, and others. Unconventional myosins have been implicated in a large number of intracellular activities, including cell growth and development, cell movement, organelle movement, protein localization and transport, endo- and exocytosis, phototransduction, hair cell adaptation, signal transduction, and others. *See, e.g.*, Mermall *et al.*, (1998) *Science* 279:527-533.

Thus, there is a need in the art for new approaches to understanding and modulating estrogen signaling in animals. The present invention addresses this and other needs.

SUMMARY OF THE INVENTION

The present invention provides novel nucleic acid and protein sequences for estrogen-regulated unconventional myosin proteins, methods for using the sequences to modulate the effects of estrogen in mammalian cells, methods of detecting estrogen signaling in cells, and methods of identifying compounds capable of acting as an estrogen receptor agonist or antagonist.

In one aspect, the present invention provides an isolated nucleic acid encoding an estrogen-regulated unconventional myosin-related protein, the protein having at least one of the following characteristics: (1) comprising at least about 70% amino acid sequence similarity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4, and 6; or (2) specifically binding to polyclonal antibodies generated against a polypeptide comprising a sequence selected from the group consisting of SEQ ID NOs: 1, 4, and 6.

In one embodiment, the protein comprises at least about 70% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4, and 6. In another embodiment, the nucleic acid encodes a protein comprising an amino acid sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6. In another embodiment, the nucleic acid encodes a protein comprising an amino acid sequence

selected from the group consisting of SEQ ID NOs: 8, 9, 10, and 11. In another embodiment, the nucleic acid comprises a nucleotide sequence that is at least about 70% similar to a sequence selected from the group consisting of SEQ ID NOs: 2, 3, 5 and 7. In another embodiment, the nucleic acid sequence comprises a nucleotide sequence that is at least about 70% identical to a sequence selected from the group consisting of SEQ ID NOs: 2,3, 5 and 7. In another embodiment, the nucleic acid comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 2, 3, 5 and 7. In another embodiment, the nucleic acid hybridizes under moderately stringent wash conditions to a nucleic acid comprising the nucleotide sequence selected from the group consisting of SEQ ID NOs:2, 3, 5 and 7. In another embodiment, the nucleic acid hybridizes under stringent wash conditions to a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs:2, 3, 5 and 7. In another embodiment, the nucleic acid is from a mouse or a human.

In another aspect, the present invention provides an expression cassette comprising the nucleic acid. In another aspect, the present invention provides an isolated eukaryotic cell comprising the expression cassette.

In another aspect, the present invention provides an isolated estrogen-regulated unconventional myosin-related protein, the protein having at least one of the following characteristics: (1) comprising at least about 70% amino acid sequence similarity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6; or (2) specifically binding to polyclonal antibodies generated against a polypeptide comprising a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6.

In one embodiment, the protein comprises at least about 70% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4 or 6. In another embodiment, the protein comprises an amino acid sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6. In another embodiment, the protein comprises an amino acid sequence selected from the group consisting of SEQ ID NO:8, 9, 10, or 11. In another embodiment, the protein is from a mouse or a human.

In another aspect, the present invention provides antibodies that selectively bind to an estrogen-regulated unconventional myosin-related protein having at least one of the following characteristics: (1) comprising at least about 70% amino acid sequence similarity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6; or (2) specifically binding to polyclonal antibodies generated against a polypeptide comprising a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6.

In one embodiment, the protein comprises at least about 70% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6.

In another aspect, the present invention provides a method of modulating the effects of estrogen in a mammalian cell, the method comprising modulating the level of expression or activity of an estrogen-regulated unconventional myosin-related protein having at least one of the following characteristics: (1) comprising at least about 70% amino acid sequence similarity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6; or (2) specifically binding to polyclonal antibodies generated against a polypeptide comprising a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6.

In one embodiment, the protein comprises at least about 70% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6. In another embodiment, the level of expression is modulated by introducing a polynucleotide into the cell, whereby the presence or expression of the polynucleotide modulates the level of expression of the estrogen-regulated unconventional myosin related protein in the cell. In another embodiment, the polynucleotide encodes a full length estrogen-regulated unconventional myosin-related protein, wherein expression of the polynucleotide increases the level of expression of the estrogen-regulated unconventional myosin-related protein in the cell. In another embodiment, the polynucleotide is an antisense sequence, wherein the presence or expression of the polynucleotide decreases the level of expression of estrogen-regulated unconventional myosin-related protein in the cell. In another embodiment, a compound is administered to the cell, whereby the level of the expression or activity of the estrogen-regulated unconventional myosin-related protein is modulated. In another embodiment, the effects of estrogen are mediated by an estrogen receptor α.

In another embodiment, the cell is present in a mammal. In another embodiment, the level of expression or activity of the estrogen-regulated unconventional myosin-related protein is increased, whereby the development of atherosclerosis or osteoporosis in the mammal is inhibited. In another embodiment, the level of expression or activity of the estrogen-regulated unconventional myosin-related protein is decreased, whereby the development of breast or other types of cancer is inhibited.

In another aspect, the present invention provides a method of detecting the presence of estrogen signaling in a mammalian cell, the method comprising detecting the expression of an estrogen-regulated unconventional myosin-related protein encoding nucleic acid in the cell, the protein: (1) comprising at least about 70% amino acid sequence similarity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6; or (2) specifically binding to polyclonal antibodies generated against a polypeptide comprising a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6.

In one embodiment, the protein comprises at least about 70% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4, and 6. In another embodiment, the presence of estrogen signaling is used in order to determine the tissue-specific distribution of estrogen signaling in a mammal. In another embodiment, the estrogen signaling is mediated by an estrogen receptor α. In another embodiment, the expression of the nucleic acid in the cell is detected by detecting the expression or activity of the encoded estrogen-regulated unconventional myosin-related protein in the cell. In another embodiment, the expression of the nucleic acid in the cell is detected by detecting the level of estrogen-regulated unconventional-myosin related protein-encoding mRNA in the cell.

In another aspect, the present invention provides a method of identifying a compound capable of acting as an estrogen-receptor agonist or antagonist, the method comprising: (1) contacting a cell comprising an estrogen receptor with the compound; and (2) detecting the functional effect of the compound on the cell, wherein an increase in the level of estrogen regulated unconventional myosin-related mRNA, protein, or protein activity in the cell indicates that the compound is capable of acting as an estrogen receptor agonist, and a decrease in the level of estrogen regulated unconventional myosin related mRNA, protein, or protein activity in the cell indicates that the compound is capable of acting as an estrogen receptor antagonist.

In one embodiment, the estrogen receptor is an estrogen receptor α. In another embodiment, the mRNA comprises at least about 70% nucleotide sequence similarity to a sequence selected from the group consisting of SEQ ID NOs: 2, 3, 5 and 7. In another embodiment, the mRNA comprises at least about 70% nucleotide sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 2, 3, 5 and 7. In another embodiment, the protein comprises at least about 70% amino acid sequence similarity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6.

In another embodiment, the protein comprises at least about 70% amino acid sequence identity to a sequence selected from the group consisting of SEQ ID NOs: 1, 4 and 6.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides the amino acid sequence for mouse MRP (mMRP).

Figure 2 provides the nucleotide sequence for a cDNA encoding mouse MRP variant 1.

Figure 3 provides the nucleotide sequence for a cDNA encoding mouse MRP variant 2.

Figure 4 provides partial amino acid sequence for human MRP1 (hMRP1).

Figure 5 provides partial nucleotide sequence for human MRP1.

Figure 6 provides partial amino acid sequence for human MRP2 (hMRP2).

Figure 7 provides partial nucleotide sequence for human MRP2.

Figure 8 provides an alignment between the amino acid sequences for mouse and human MRP.

Figure 9 illustrates the domain structure for MRP. The domain structure was analyzed with HMM using the translated amino acid sequence.

Figure 10 illustrates the regulation of the MRP gene by estrogen. Northern hybridization of liver RNA from wildtype (WT) mice and knock-out mice (ERKO), which lack the estrogen receptor α, all of which were treated with vehicle (V), 17β-estradiol (E2) and antiestrogen compound (ZK). The myosin-related protein (MRP) gene was up-regulated by E2 in the WT mice only, demonstrating that this gene is regulated via the estrogen receptor α in liver.

Figure 11 illustrates the tissue specific regulation of mMRP gene by estrogen. RT-PCR was performed on total RNA from WT or ERKO liver and brain tissues treated with vehicle (V), 17β-estradiol (E2) and antiestrogen compound (ZK). RNA quantity was controlled by RT-PCR on a house-keeping gene (GAPDH) in the same experiments. The MRP gene was differentially regulated, *i.e.*, it is regulated by the estrogen receptor α in liver and by the estrogen receptor β in brain.

Figure 12 illustrates the chromosomal location of mMRP. mMRP has has been localized to chromosome 11.

Figure 13 illustrates the chromosomal location of hMRP. hMRP has has been localized to chromosome 17 and, more specifically, to 17q25.

DETAILED DESCRIPTION OF THE INVENTION AND PREFERRED EMBODIMENTS

I. Introduction

The present invention provides nucleic acids and polypeptides for MRP, a novel estrogen-regulated unconventional myosin-related protein. MRP is dramatically upregulated by estrogen, indicating that MRP is a molecular mediator of the *in vivo* function of estrogen. In fact, it has been discovered that the MRP gene is differentially regulated, *i.e.*, it is regulated by the estrogen receptor α in liver and by the estrogen receptor β in brain (*see*, Figures 10-11). Accordingly, the effects of estrogen can be modulated in cells by modulating the expression or activity of MRP, and, in addition, the presence, level, or tissue distribution of estrogen signaling can be detected by detecting MRP expression or activity. MRP can also be used as a marker gene for the investigation of tissue-specific and estrogen receptor-specific ligands, such as estrogen receptor agonists and antagonists.

Numerous MRP sequences are provided. For example, two variants of a mouse MRP (mMRP) cDNA, shown as SEQ ID NOs: 2 and 3, are provided. These cDNAs encode a single protein, shown as SEQ ID NO:1. In addition, two human MRP cDNA sequences (hMRP1 and hMRP2) are shown as SEQ ID NOs: 5 and 7, respectively. These cDNAs encode two distinct proteins, shown as SEQ ID NOs: 4 and 6. In addition, signature sequences that are conserved between the mouse and human MRP genes are provided (SEQ ID NOs:8-11). These signature sequences can be used to identify novel MRP genes from mouse, humans, or any species. The genomic position of human (17q25) and mouse (chromosome 11) MRP is also provided (*see*, Figures 12 and 13).

Modulators, recombinant forms, derivatives, variants, or fragments of the herein-described MRPs can be used to enhance or inhibit estrogen signaling in cells, and can therefore be useful for the treatment of a wide variety of estrogen-related diseases. For example, estrogen signaling can be enhanced to treat diseases including, but not limited to, osteoporosis, cardiovascular diseases, Alzheimer's Disease, or Parkinson's Disease. Alternatively, estrogen signaling can be inhibited to treat, *e.g.*, breast or other cancers.

Moreover, it has been discovered that MRP maps to chromosome 17 in humans (chromosome 11 in mice), a chromosome that has previously been identified to be important in hereditary deafness. In fact, it has previously been discovered that mutations in genes encoding other uncoventional myosin proteins (*e.g.*, *MYO*15, *Shaker–*

2, *etc.*) cause deafness and circuling behavior (*see*, Probst, *et al.*, *Science*, 280:1444-1447 (May 1998); and Wang, *et al.*, *Science,* 280:1447-1450 (May 1998), the teachings of both of which are incorporated herein by reference for all purposes). It has now been discovered that a mutation(s) in the MRP gene of the present invention results in hearing loss and/or deafness. As such, modulation of the MRP gene can be used to prevent and/or to treat hearing loss and deafness. Moreover, ERKO mice and/or BERKO mice are good models for examing the role of MRP in the auditory system (*e.g.*, the auditory signaling pathways, auditory hair cell function, *etc.*).

In numerous embodiments, the present invention provides methods of screening for modulators, *e.g.*, activators, inhibitors, stimulators, enhancers, *etc.*, of MRP nucleic acids and proteins. Such modulators can affect MRP activity in any of a number of ways, *e.g.*, by modulating MRP transcription, translation, phosphorylation, RNA or protein stability, by altering the binding of MRP to heterologous proteins or other molecules; or by affecting MRP protein activity. In preferred embodiments, modulators that increase or decrease MRP activity or levels are used to treat any of the above-recited estrogen-related diseases or conditions.

In one embodiment, compounds are screened, *e.g.*, using high throughput screening (HTS), to identify those compounds that can bind to and/or modulate the activity of an isolated MRP polypeptide or fragment thereof. In another embodiment, MRP proteins are recombinantly expressed in cells, and potential modulators of MRP are assayed by detecting the presence or activity of the MRP in the cells.

In numerous embodiments, an MRP polynucleotide or polypeptide is introduced into a cell, *in vivo* or *ex vivo*, and the MRP activity in the cell is thereby modulated. For example, a polynucleotide encoding a full length MRP polypeptide is introduced into a population of cells, thereby increasing the level or activity of MRP in the cells. Alternatively, an antisense, ribozyme, or dominant-negative encoding polynucleotide can be introduced into a population of cells, thereby inhibiting the MRP levels and/or activity in the cells.

The present invention also provides methods for detecting MRP nucleic acid and protein expression, allowing investigation into estrogen-mediated signaling and transcription, *e.g.*, through an estrogen receptor α or, alternatively, through an estrogen receptor β, and allowing the specific identification, *in vitro* or *in vivo,* of estrogen responsive cells. MRP also provides useful nucleic acid probes for paternity and forensic

investigations. MRP polypeptides can also be used to generate monoclonal and polyclonal antibodies useful for identifying estrogen responsive cells, particularly cells responsive to estrogen through an estrogen receptor α or, alternatively, through an estrogen receptor β. MRP expression can be identified using techniques such as reverse transcription and amplification of mRNA, isolation of total RNA or poly A^+ RNA, northern blotting, dot blotting, *in situ* hybridization, RNase protection, S1 digestion, probing DNA microchip arrays, western blots, and the like.

Functionally, MRP nucleic acids encode novel members of the unconventional myosin-related protein family that are strongly upregulated by estrogen signaling. MRP proteins comprise several functionally important domains, including, *e.g.*, a myosin head, an SH3 domain, IQ domains, a talin domain, and an MyTH4 domain (*see*, Chen *et al.*, (1996) *Genomics* 36:440-448) (*see*, *e.g.*, Figure 9). SH3 domains are typically involved in protein-protein interactions, and are present in a number of signaling molecules. Related MRP genes from other species share at least about 60% nucleotide sequence similarity or identity over a region of at least about 50 nucleotides in length, optionally 100, 200, 500, or more nucleotides in length, to SEQ ID NO:2, 3, 5 or 7or encode polypeptides sharing at least about 60% amino acid sequence similarity or identity over an amino acid region at least about 25 amino acids in length, optionally 50 to 100 amino acids in length to SEQ ID NO:1, 4 or 6. In particular, related MRP genes can be identified by examining the similarity of a candidate MRP to a region conserved between mMRP and hMRP, *e.g.*, a conserved region as shown in Figure 8 (*see*, *e.g.*, SEQ ID NOs:8-11). Preferably, the MRP polypeptide comprises from about 1200-1400 amino acids, *e.g.*, about 1337 amino acids.

The present invention also provides polymorphic variants of the mMRP protein depicted in SEQ ID NO:1: variant #1, in which a leucine residue is substituted for an isoleucine residue at amino acid position 45; and variant #2, in which an alanine residue is substituted for a glycine residue at amino acid position 30.

The present invention also provides polymorphic variants of the hMRP1 protein depicted in SEQ ID NO:4: variant #1, in which an isoleucine residue is substituted for a leucine residue at amino acid position 23; and variant #2, in which an arginine residue is substituted for a lysine residue at amino acid position 29.

The present invention also provides polymorphic variants of the hMRP2 protein depicted in SEQ ID NO:6: variant #1, in which a lysine residue is substituted for

an arginine residue at amino acid position 63; and variant #2, in which a valine residue is substituted for a leucine residue at amino acid position 70.

Specific regions of the MRP nucleotide and amino acid sequences may be used to identify polymorphic variants, interspecies homologs, and alleles of MRP genes. This identification can be made *in vitro*, *e.g.*, under stringent hybridization conditions, or PCR and sequencing, or by using the sequence information in a computer system for comparison with other nucleotide sequences. Typically, identification of polymorphic variants and alleles of an MRP protein is made by comparing an amino acid sequence of about 25 amino acids or more, *e.g.*, 50-100 amino acids. Such regions preferably encompass a region highly conserved among MRP proteins, *e.g.*, a conserved region as shown in Figure 8 such as the signatures sequences shown in SEQ ID NOs:8-11. Amino acid identity of approximately at least 60% or above, optionally 65%, 70%, 75%, 80%, 85%, or 90-95% or above typically demonstrates that a protein is a polymorphic variant, interspecies homolog, or allele of the MRP protein. Sequence comparison can be performed using any of the sequence comparison algorithms discussed below. Antibodies that bind specifically to MRP polypeptides or a conserved region thereof can also be used to identify alleles, interspecies homologs, and polymorphic variants.

Polymorphic variants, interspecies homologs, and alleles of the herein-described MRPs can be confirmed functionally, *e.g.*, by examining the expression or activity of the putative homolog in response to estrogen. Typically, an MRP polypeptide comprising an amino acid sequence of SEQ ID NO:1, 4 or 6 is used as a positive control in comparison to the putative MRP protein to demonstrate the identification of a polymorphic variant or allele of the MRP gene or protein.

Nucleotide and amino acid sequence information for the present MRP sequences may also be used to construct models of MRP polypeptides in a computer system. These models are subsequently used to identify compounds that can activate or inhibit MRP proteins. Such compounds can be used to investigate the role of MRP genes in estrogen mediated signaling, and to treat any of a number of estrogen related diseases or conditions.

The present invention also provides assays, preferably high throughput assays, to identify compounds or other molecules that interact with and/or modulate an MRP. In certain assays, a particular domain of the MRP is used, *e.g.*, a myosin head domain, an SH3 domain, an MyTH4 domain, or an IQ motif.

The present invention also provides methods to treat diseases or conditions associated with estrogen signaling. For example, MRP activity and/or expression can be altered in cells of a patient to treat or prevent diseases and conditions including, but not limited to, atherosclerosis, osteoporosis, Alzheimer's Disease, Parkinson's Disease, and breast cancer.

Transgenic animals and cells lacking one or more MRP alleles, or containing altered forms of an MRP are also provided, as are kits for using the herein-disclosed polynucleotides and polypeptides and for practicing the herein-disclosed methods.

II. Definitions

As used herein, the following terms have the meanings ascribed to them below unless specified otherwise.

As used herein, "MRP" refers to a unconventional myosin-related protein as shown in SEQ ID NO:1, 4 or 6, or any derivative, homolog, or fragment thereof, or to any nucleic acid encoding such a protein, derivative, homolog, or fragment thereof. MRP proteins or derivatives can be expressed in any cell type, including any eukaryotic or prokaryotic cell, or synthesized *in vitro*. It will be recognized that derivatives, homologs, and fragments of MRP can readily be used in the present invention. Such MRP variants can comprise any one or more domains of the polypeptide shown as SEQ ID NO:1, 4 or 6, or multiple copies of any one or more domains, or any number of domains in novel combinations with each other or with other proteins or protein domains. The term "MRP" also refers to polymorphic variants, alleles, mutants, and interspecies homologs that: (1) have about 60% amino acid sequence similarity or identity, optionally about 75, 80, 85, 90, or 95% amino acid sequence similarity or identity to SEQ ID NO:1, 4 or 6 over a window of about 25 amino acids, optionally 50-100 amino acids; (2) specifically bind to antibodies raised against an immunogen comprising an amino acid sequence of SEQ ID NO:1, 4 or 6, and conservatively modified variants thereof; (3) specifically hybridize (with a size of at least about 100, optionally at least about 500-1000 nucleotides) under stringent hybridization conditions to a sequence of SEQ ID NO:2, 3, 5, 7, and conservatively modified variants thereof; or (4) comprise an MRP signature sequence as shown in SEQ ID NO:8-11, or any conserved sequence, *e.g.*, as shown in Figure 8.. "MRP" polynucleotides or polypeptides can be derived from any eukaryotic source, including mammals such as mice or humans.

"mMRP" refers to an MRP protein derived from mouse, as shown in SEQ ID NO:1, or to derivatives, fragments, or variants thereof, or to mouse-derived MRP-encoding nucleic acids, as shown in SEQ ID NOs: 2 and 3, or to derivatives, fragments, or variants thereof.

"hMRP" refers to an MRP protein derived from human, as shown in SEQ ID NOs: 4 or 6, or to derivatives, fragments, or variants thereof, or to human-derived MRP-encoding nucleic acids, as shown in SEQ ID NOs: 5 and 7, or to derivatives, fragments, or variants thereof.

Topologically, full-length mMRP polypeptides include, *inter alia*, a head domain, a neck domain, two IQ motifs, a talin domains, a tail domain, an SH3 domain, and a MyTH4 domain. These domains can be structurally identified using methods known to those of skill in the art, such as standard sequence analysis programs and by comparison with related proteins.

The head domain refers to a region corresponding to, *e.g.*, approximately the 700 N-terminal amino acids of SEQ ID NO:1, and which is homologous to a region of other members of the myosin superfamily. Often, myosin head domains contain ATP-binding and actin-binding domains, but these domains are absent in mMRP.

The "MyTH4 domain" refers to a conserved region of the mMRP protein that is homologous to a domain found within the tail domain of four divergent members of the myosin superfamily (*see, e.g.*, Chen *et al.*, (1996) *Genomics* 36:440-448). In mMRP, the "MyTH4 domain" extends approximately from amino acid 346 to amino acid 458 of SEQ ID NO:1.

The "SH3 domain," or "Src-homology domain 3," is often involved in mediating protein-protein interactions, and is present in a number of signaling molecules. In mMRP, the "SH3 domain" corresponds approximately to amino acids 1247-1303.

IQ motifs refer to light-chain binding motifs that are common to members of the myosin superfamily. The IQ domains in, *e.g.*, mMRP, correspond approximately to amino acid positions 161-171 and 180-190 of SEQ ID NO:1. *See, e.g.*, Wang *et al.*, (1998) *Science* 280:1447-1451.

"Biological sample," as used herein, refers to a sample of biological tissue or fluid that contains one or more MRP nucleic acids encoding one or more MRP proteins. Such samples include, but are not limited to, tissue isolated from humans and mice, in particular, liver, brain, lung, and kidneys. Biological samples may also include sections of tissues such as frozen sections taken for histological purposes. A biological

sample is typically obtained from a eukaryotic organism, such as insects, protozoa, birds, fish, reptiles, and preferably a mammal such as rat, mouse, cow, dog, guinea pig, or rabbit, and most preferably a primate such as a chimpanzee or a human.

By "determining the functional effect" is meant assaying for a compound that modulates, *e.g.*, increases or decreases, a parameter that is indirectly or directly under the influence of an MRP polynucleotide or polypeptide, *e.g.*, functional, physical and chemical effects. Such functional effects can be measured by any means known to those skilled in the art, *e.g.*, changes in spectroscopic characteristics (*e.g.*, fluorescence, absorbance, refractive index), hydrodynamic (*e.g.*, shape), chromatographic, or solubility properties, changes in the level of MRP mRNA, protein, or protein activity, or of any cellular effect indicative of MRP activity.

"Inhibitors," "activators," and "modulators" of MRP genes or proteins are used to refer to inhibiting, activating, or modulating molecules identified using *in vitro* and *in vivo* assays for MRP. Inhibitors are compounds that, *e.g.*, bind to MRP proteins, partially or totally block MRP activity, downregulate MRP expression or stability, or prevent MRP binding to heterologous molecules. Activators are compounds that, *e.g.*, bind to MRP, stimulate MRP activity, increase MRP expression or stability, or facilitate MRP binding to membranes or to other proteins or factors. Modulators may include genetically modified versions of MRP proteins, *e.g.*, dominant negative or activated forms of MRP. Such assays for inhibitors and activators are described below and include, *e.g.*, expressing MRP proteins in cells, applying putative modulator compounds, and then determining the functional effects on MRP activity. Samples or assays comprising MRP polypeptides that are treated with a potential activator, inhibitor, or modulator are compared to control samples without the inhibitor, activator, or modulator to examine the effect of the candidate compound. Control samples (untreated with the compound) are assigned a relative MRP activity value of 100%. Inhibition of an MRP polypeptide is achieved when the activity value relative to the control is about 80%, optionally 50% or 25-0%. Activation of an MRP polypeptide is achieved when the activity value relative to the control is 110%, optionally 150%, optionally 200-500%, or 1000-3000% higher.

The terms "isolated" "purified" or "biologically pure" refer to material that is substantially or essentially free from components which normally accompany it as found in its native state. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein that is the predominant species present in

a preparation is substantially purified. In particular, an isolated MRP nucleic acid is separated from open reading frames that flank the MRP gene and which encode proteins other than MRP. The term "purified" denotes that a nucleic acid or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the nucleic acid or protein is at least 85% pure, optionally at least 95% pure, and optionally at least 99% pure.

"Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs).

Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al., Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al., J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al., Mol. Cell. Probes* 8:91-98 (1994)). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide, and polynucleotide.

The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers.

The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino

acid, *i.e.*, an α carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refer to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

"Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or, where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids can encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well

known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The following eight groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Glycine (G);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V);
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W);
7) Serine (S), Threonine (T); and
8) Cysteine (C), Methionine (M)

(*see, e.g.*, Creighton, *Proteins* (1984)).

Macromolecular structures such as polypeptide structures can be described in terms of various levels of organization. For a general discussion of this organization, *see, e.g.*, Alberts *et al.*, *Molecular Biology of the Cell* (3rd ed., 1994) and Cantor and Schimmel, *Biophysical Chemistry Part I: The Conformation of Biological Macromolecules* (1980). "Primary structure" refers to the amino acid sequence of a particular peptide. "Secondary structure" refers to locally ordered, three-dimensional structures within a polypeptide. These structures are commonly known as domains. Domains are portions of a polypeptide that form a compact unit of the polypeptide and are typically 50 to 350 amino acids long. Typical domains are made up of sections of lesser organization such as stretches of β-sheet and α-helices. "Tertiary structure" refers to the complete three dimensional structure of a polypeptide monomer. "Quaternary structure" refers to the three dimensional structure formed by the noncovalent association of independent tertiary units. Anisotropic terms are also known as energy terms.

A "label" or a "detectable moiety" is a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ^{32}P, fluorescent dyes, electron-dense reagents, enzymes (*e.g.*, as commonly used in an ELISA), biotin, digoxigenin, or haptens and proteins which can be made detectable, *e.g.*, by incorporating a radiolabel into the peptide or used to detect antibodies specifically reactive with the peptide.

A "labeled nucleic acid probe or oligonucleotide" is one that is bound, either covalently, through a linker or a chemical bond, or noncovalently, through ionic, van der Waals, electrostatic, or hydrogen bonds to a label such that the presence of the probe may be detected by detecting the presence of the label bound to the probe.

As used herein a "nucleic acid probe or oligonucleotide" is defined as a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation. As used herein, a probe may include natural (*i.e.*, A, G, C, or T) or modified bases (7-deazaguanosine, inosine, *etc.*). In addition, the bases in a probe may be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, for example, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages. It will be understood by one of skill in the art that probes may bind target sequences lacking complete complementarity with the probe sequence depending upon the stringency of the hybridization conditions. The probes are optionally directly labeled as with isotopes, chromophores, lumiphores, chromogens, or indirectly labeled such as with biotin to which a streptavidin complex may later bind. By assaying for the presence or absence of the probe, one can detect the presence or absence of the select sequence or subsequence.

The term "recombinant" when used with reference, *e.g.*, to a cell, or nucleic acid, protein, or vector, indicates that the cell, nucleic acid, protein or vector, has been modified by the introduction of a heterologous nucleic acid or protein or the alteration of a native nucleic acid or protein, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found within the native (nonrecombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all.

The term "heterologous" when used with reference to portions of a nucleic acid indicates that the nucleic acid comprises two or more subsequences that are not found in the same relationship to each other in nature. For instance, the nucleic acid is typically recombinantly produced, having two or more sequences from unrelated genes arranged to make a new functional nucleic acid, *e.g.*, a promoter from one source and a coding region from another source. Similarly, a heterologous protein indicates that the protein comprises two or more subsequences that are not found in the same relationship to each other in nature (*e.g.*, a fusion protein).

A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of a nucleic acid. As used herein, a promoter includes necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. A "constitutive" promoter is a promoter that is active under most environmental and developmental conditions. An "inducible" promoter is a promoter that is active under environmental or developmental regulation. The term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

An "expression vector" is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a nucleic acid to be transcribed operably linked to a promoter.

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.*, 60% identity, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% identity over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the compliment of a test sequence. Optionally, the identity exists over a region that is at least about 50 amino acids or nucleotides in length, or more preferably over a region that is at least about 75-100 amino acids or nucleotides in length.

The term "similarity," or percent "similarity," in the context of two or more polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of amino acid residues that are either the same or similar as defined in the 8 conservative amino acid substitutions defined above (*i.e.*, 60%, optionally 65%, 70%, 75%, 80%, 85%, 90%, or 95% similar over a specified region), when compared and aligned for maximum correspondence over a comparison window, or designated region as

measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially similar." Optionally, this similarity exists over a region that is at least about 50 amino acids in length, or more preferably over a region that is at least about 75-100 amino acids in length.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities or similarity for the test sequences relative to the reference sequence, based on the program parameters.

A "comparison window," as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and visual inspection (*see, e.g., Current Protocols in Molecular Biology* (Ausubel *et al.*, eds. 1995 supplement)).

One example of a useful algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments to show relationship and percent sequence identity. It also plots a tree or dendogram showing the clustering relationships used to create the alignment. PILEUP uses a simplification of the progressive alignment method of Feng & Doolittle, *J. Mol. Evol.* 35:351-360 (1987). The method used is similar to the method described by Higgins & Sharp, *CABIOS* 5:151-153 (1989). The program can align up to 300 sequences, each

of a maximum length of 5,000 nucleotides or amino acids. The multiple alignment procedure begins with the pairwise alignment of the two most similar sequences, producing a cluster of two aligned sequences. This cluster is then aligned to the next most related sequence or cluster of aligned sequences. Two clusters of sequences are aligned by a simple extension of the pairwise alignment of two individual sequences. The final alignment is achieved by a series of progressive, pairwise alignments. The program is run by designating specific sequences and their amino acid or nucleotide coordinates for regions of sequence comparison and by designating the program parameters. Using PILEUP, a reference sequence is compared to other test sequences to determine the percent sequence identity relationship using the following parameters: default gap weight (3.00), default gap length weight (0.10), and weighted end gaps. PILEUP can be obtained from the GCG sequence analysis software package, *e.g.*, version 7.0 (Devereaux *et al.*, *Nuc. Acids Res.* 12:387-395 (1984)).

Another example of an algorithm that is suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al., Nuc. Acids Res.* 25:3389-3402 (1977) and Altschul *et al., J. Mol. Biol.* 215:403-410 (1990), respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences)

uses as defaults a wordlength (W) of 11, an expectation (E) or 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands.

The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Natl. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (*e.g.*, total cellular or library DNA or RNA).

The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic*

Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10° C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60° C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, optionally 10 times background hybridization. Exemplary stringent hybridization conditions can be as following: 50% formamide, 5x SSC, and 1% SDS, incubating at 42°C, or, 5x SSC, 1% SDS, incubating at 65°C, with wash in 0.2x SSC, and 0.1% SDS at 65°C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes.

Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cases, the nucleic acids typically hybridize under moderately stringent hybridization conditions. Exemplary "moderately stringent hybridization conditions" include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 1X SSC at 45°C. Such washes can be performed for 5, 15, 30, 60, 120, or more minutes. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

"Antibody" refers to a polypeptide comprising a framework region from an immunoglobulin gene or fragments thereof that specifically binds and recognizes an antigen. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon, and mu constant region genes, as well as the myriad immunoglobulin variable region genes. Light chains are classified as either kappa or

lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kDa) and one "heavy" chain (about 50-70 kDa). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively.

Antibodies exist, *e.g.*, as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce F(ab)'$_2$, a dimer of Fab which itself is a light chain joined to V_H-C_H1 by a disulfide bond. The F(ab)'$_2$ may be reduced under mild conditions to break the disulfide linkage in the hinge region, thereby converting the F(ab)'$_2$ dimer into an Fab' monomer. The Fab' monomer is essentially Fab with part of the hinge region (*see Fundamental Immunology* (Paul ed., 3d ed. 1993). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by using recombinant DNA methodology. Thus, the term antibody, as used herein, also includes antibody fragments either produced by the modification of whole antibodies, or those synthesized *de novo* using recombinant DNA methodologies (*e.g.*, single chain Fv) or those identified using phage display libraries (*see, e.g.*, McCafferty *et al.*, *Nature* 348:552-554 (1990)).

For preparation of monoclonal or polyclonal antibodies, any technique known in the art can be used (*see, e.g.*, Kohler & Milstein, *Nature* 256:495-497 (1975); Kozbor *et al.*, *Immunology Today* 4: 72 (1983); Cole *et al.*, pp. 77-96 in *Monoclonal Antibodies and Cancer Therapy* (1985)). Techniques for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies. Alternatively, phage display technology can be used to identify antibodies and heteromeric Fab fragments that specifically bind to selected antigens (*see, e.g.*, McCafferty *et al.*, *Nature* 348:552-554 (1990); Marks *et al.*, *Biotechnology* 10:779-783 (1992)).

A "chimeric antibody" is an antibody molecule in which (a) the constant region, or a portion thereof, is altered, replaced or exchanged so that the antigen binding

site (variable region) is linked to a constant region of a different or altered class, effector function and/or species, or an entirely different molecule which confers new properties to the chimeric antibody, *e.g.*, an enzyme, toxin, hormone, growth factor, drug, etc.; or (b) the variable region, or a portion thereof, is altered, replaced or exchanged with a variable region having a different or altered antigen specificity.

An "anti-MRP" antibody is an antibody or antibody fragment that specifically binds a polypeptide encoded by an MRP gene, cDNA, or a subsequence thereof.

The term "immunoassay" is an assay that uses an antibody to specifically bind an antigen. The immunoassay is characterized by the use of specific binding properties of a particular antibody to isolate, target, and/or quantify the antigen.

The phrase "specifically (or selectively) binds" to an antibody or "specifically (or selectively) immunoreactive with," when referring to a protein or peptide, refers to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and do not substantially bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, polyclonal antibodies raised to an MRP polypeptide from specific species such as rat, mouse, or human can be selected to obtain only those polyclonal antibodies that are specifically immunoreactive with the MRP protein and not with other proteins, except for polymorphic variants and alleles of the MRP protein. This selection may be achieved by subtracting out antibodies that cross-react with MRP molecules from other species. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (*see, e.g.*, Harlow & Lane, *Antibodies, A Laboratory Manual* (1988), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

The phrase "selectively associates with" refers to the ability of a nucleic acid to "selectively hybridize" with another as defined above, or the ability of an antibody to "selectively (or specifically) bind to a protein, as defined above.

By "host cell" is meant a cell that contains an expression vector and supports the replication or expression of the expression vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells such as CHO, HeLa and the like, *e.g.*, cultured cells, explants, and cells *in vivo*.

III. Manipulation and detection of MRP nucleic acids

In numerous embodiments of the present invention, nucleic acids encoding an MRP polypeptide, including full-length MRP proteins, or any derivative, variant, homolog, or fragment thereof, will be used. Such nucleic acids are useful for any of a number of applications, including for the production of MRP proteins, for diagnostic assays, for therapeutic applications, for use as MRP-specific probes, for assays for MRP-binding and/or modulating compounds, to identify and/or isolate MRP homologs from other species, and for other applications.

A. General recombinant DNA methods

Numerous applications of the present invention involve the cloning, synthesis, maintenance, mutagenesis, and other manipulations of nucleic acid sequences that can be performed using routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)).

For nucleic acids, sizes are given in either kilobases (kb) or base pairs (bp). These are estimates derived from agarose or acrylamide gel electrophoresis, from sequenced nucleic acids, or from published DNA sequences. For proteins, sizes are given in kilodaltons (kDa) or amino acid residue numbers. Protein sizes are estimated from gel electrophoresis, from sequenced proteins, from derived amino acid sequences, or from published protein sequences.

Oligonucleotides that are not commercially available can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage & Caruthers, *Tetrahedron Letts.* 22:1859-1862 (1981), using an automated synthesizer, as described in Van Devanter *et al.*, *Nucleic Acids Res.* 12:6159-6168 (1984). Purification of oligonucleotides is by either native acrylamide gel electrophoresis or by

anion-exchange HPLC as described in Pearson & Reanier, *J. Chrom.* 255:137-149 (1983).

The sequence of the cloned genes and synthetic oligonucleotides can be verified after cloning using, *e.g.*, the chain termination method for sequencing double-stranded templates of Wallace *et al., Gene* 16:21-26 (1981).

B. Isolating and Detecting MRP nucleotide sequences

In numerous embodiments of the present invention, MRP nucleic acids are isolated and cloned using recombinant methods. Such embodiments are used, *e.g.*, to isolate MRP polynucleotides for protein expression or during the generation of variants, derivatives, expression cassettes, or other sequences derived from MRP, to monitor MRP gene expression, for the isolation or detection of MRP sequences in different species, for diagnostic purposes in a patient, *i.e.,* to detect mutations in MRP, or for genotyping and/or forensic applications.

Often, the nucleic acid sequences encoding MRP proteins and related nucleic acid sequence homologs are cloned from cDNA and genomic DNA libraries by hybridization with probes, or isolated using amplification techniques with oligonucleotide primers. For example, MRP sequences are typically isolated from mammalian nucleic acid (genomic or cDNA) libraries by hybridizing with a nucleic acid probe, the sequence of which can be derived from, *e.g.*, SEQ ID NO: 2, 3, 5 or 7 or amplified using primers designed using SEQ ID NO: 2, 3, 5 or 7. A suitable biological material from which RNA and cDNA for MRP can be isolated is, *e.g.*, brain, liver, kidney, or lung. One especially effective method to identify novel MRPs, *e.g.*, an MRP from a species other than human or mouse, is to use primers or probes from sequences encoding especially conserved regions of an MRP, *e.g.*, a region that is identical between an mMRP and an hMRP as shown in Figure 8 (*see, e.g.*, SEQ ID NOs:8-11).

Amplification techniques using primers can also be used to amplify and isolate MRP sequences from DNA or RNA (*see, e.g.*, Dieffenfach & Dveksler, *PCR Primer: A Laboratory Manual* (1995)). Primers can be used, *e.g.*, to amplify either the full length sequence or a probe of from one to several hundred nucleotides, which is then used to screen a mammalian library for full-length MRP clones.

Nucleic acids encoding MRP polypeptides can also be isolated from expression libraries using antibodies as probes. Such polyclonal or monoclonal

antibodies can be raised using the sequence of, *e.g.*, SEQ ID NO:1, 4 or 6, or derivatives or fragments thereof.

Polymorphic variants, alleles, and interspecies homologs that are substantially identical to an MRP gene can be isolated using MRP nucleic acid probes, and oligonucleotides under stringent hybridization conditions, by screening libraries. Alternatively, expression libraries can be used to clone MRP polymorphic variants, alleles, and interspecies homologs, by detecting expressed homologs immunologically with antisera or purified antibodies made against an MRP polypeptide, which also recognize and selectively bind to the MRP homolog.

More distantly related MRP homologs can be identified using any of a number of well known techniques, including by hybridizing an MRP probe with a genomic or cDNA library using moderately stringent conditions, or under low stringency conditions. Also, a distant homolog can be amplified from a nucleic acid library using degenerate primer sets, *i.e.*, primers that incorporate all possible codons encoding a given amino acid sequence, in particular based on a highly conserved amino acid stretch. Such primers are well known by those of skill in the art, and numerous programs are available, *e.g.*, on the Internet, for degenerate primer design.

To make a cDNA library, one should choose a source that is rich in MRP mRNA, *e.g.*, cells isolated from brain, kidney, liver or lung. The mRNA is then made into cDNA using reverse transcriptase, ligated into a recombinant vector, and transfected into a recombinant host for propagation, screening and cloning. Methods for making and screening cDNA libraries are well known (*see, e.g.,* Gubler & Hoffman, *Gene* 25:263-269 (1983); Sambrook *et al., supra*; Ausubel *et al., supra*).

For a genomic library, the DNA is extracted from the tissue or cells and either mechanically sheared or enzymatically digested to yield fragments of about 12-20 kb. The fragments are then separated by gradient centrifugation from undesired sizes and are constructed in bacteriophage lambda vectors. These vectors and phage are packaged *in vitro*. Recombinant phage are analyzed by plaque hybridization as described in Benton & Davis, *Science* 196:180-182 (1977). Colony hybridization is carried out as generally described in Grunstein *et al.*, *Proc. Natl. Acad. Sci. USA* 72:3961-3965 (1975).

An alternative method of isolating MRP nucleic acids and their homologs combines the use of synthetic oligonucleotide primers and amplification of an RNA or DNA template (*see,* U.S. Patent Nos. 4,683,195 and 4,683,202; *PCR Protocols: A Guide to Methods and Applications* (Innis *et al.*, eds, 1990)). Methods such as polymerase chain

reaction (PCR) and ligase chain reaction (LCR) can be used to amplify nucleic acid sequences of MRP genes directly from mRNA, from cDNA, from genomic libraries or cDNA libraries. Degenerate oligonucleotides can be designed to amplify MRP homologs using the sequences provided herein. Restriction endonuclease sites can be incorporated into the primers. Polymerase chain reaction or other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of MRP-encoding mRNA in physiological samples, for nucleic acid sequencing, or for other purposes. Genes amplified by the PCR reaction can be purified from agarose gels and cloned into an appropriate vector.

Synthetic oligonucleotides can be used to construct recombinant MRP genes for use as probes or for expression of protein. This method is performed using a series of overlapping oligonucleotides usually 40-120 bp in length, representing both the sense and non-sense strands of the gene. These DNA fragments are then annealed, ligated and cloned. Alternatively, amplification techniques can be used with precise primers to amplify a specific subsequence of the MRP nucleic acid. The specific subsequence is then ligated into an expression vector.

The nucleic acid encoding an MRP polypeptide is typically cloned into intermediate vectors before transformation into prokaryotic or eukaryotic cells for replication and/or expression. These intermediate vectors are typically prokaryotic vectors, *e.g.*, plasmids, or shuttle vectors. Vectors, cells, and transfection methods are well known to those of skill and are described, *e.g.*, in Ausubel or in Sambrook, both *supra*.

Optionally, nucleic acids will be used that encode chimeric proteins comprising an MRP polypeptide or domains thereof in combination with a heterologous polypeptide or polypeptides. For example, a domain such as an SH3 domain, a MyTH4 domain, a myosin head domain, an IQ motif, or a talin domain, can be covalently linked to a heterologous protein including, but not limited to, luciferase, green fluorescent protein (GFP), and β-gal, each of which is well known in the art.

In certain embodiments, MRP polynucleotides will be detected using hybridization-based methods to determine, *e.g.*, MRP RNA levels or to detect particular DNA sequences, *e.g.*, for genotyping or for forensic applications. For example, gene expression of MRP can be analyzed by techniques known in the art, *e.g.*, Northern

blotting, reverse transcription and amplification of mRNA, dot blotting, *in situ* hybridization, RNase protection, probing DNA microchip arrays, and the like. In one embodiment, high density oligonucleotide analysis technology (*e.g.*, GeneChip™) is used to identify homologs and polymorphic variants of MRP, or to monitor levels of MRP mRNA. In the case where a homolog is linked to a known disease, they can be used with GeneChip™ as a diagnostic tool in detecting the disease in a biological sample, *see, e.g.*, Gunthand *et al., AIDS Res. Hum. Retroviruses* 14: 869-876 (1998); Kozal *et al., Nat. Med.* 2:753-759 (1996); Matson *et al., Anal. Biochem.* 224:110-106 (1995); Lockhart *et al., Nat. Biotechnol.* 14:1675-1680 (1996); Gingeras *et al., Genome Res.* 8:435-448 (1998); Hacia *et al., Nucleic Acids Res.* 26:3865-3866 (1998).

MRP DNA sequences , *e.g.*, a particular MRP allele, can be detected in a mammal using Southern blot hybridization, *i.e.*, by isolating genomic DNA, performing a restriction digest on the isolated DNA, separating the restriction fragments electrophoretically, *e.g.*, in an agarose gel, and transferring the separated DNA to a membrane and probing with a specific, labeled sequence. Southern blotting is well known to those of skill in the art and is taught in numerous sources, including Ausubel *et al.* and Sambrook *et al.*

In other embodiments, *e.g.*, to detect tissue specific or temporal patterns of gene expression, *e.g.*, as a marker for estrogen signaling, an MRP polynucleotide is detected using *in situ* hybridization. In *in situ* hybridization, the target nucleic acid is liberated from its cellular surroundings in such as to be available for hybridization within the cell while preserving the cellular morphology for subsequent interpretation and analysis. The following articles provide an overview of the art of *in situ* hybridization: Singer *et al., Biotechniques* 4:230-250 (1986); Haase *et al., Methods in Virology*, vol. VII, pp. 189-226 (1984); and *Nucleic Acid Hybridization: A Practical Approach* (Hames *et al.*, eds. 1987).

C. Expression in prokaryotes and eukaryotes

Often, a cloned MRP sequence will be expressed in a prokaryotic or eukaryotic cell to obtain expression, *i.e.*, production of the encoded mRNA or protein. For example, in numerous embodiments, an MRP polynucleotide is introduced into a population of cells to modulate the level of MRP activity in the cells, thereby modulating the level of estrogen signaling in the cells, *e.g*, estrogen receptor α specific signaling or, alternatively, estrogen receptor β specific signaling. To obtain high level expression of a

cloned gene or nucleic acid, such as a cDNA encoding an MRP polypeptide, an MRP sequence is typically subcloned into an expression vector that contains a strong promoter to direct transcription, a transcription/translation terminator, and if for a nucleic acid encoding a protein, a ribosome binding site for translational initiation. Suitable bacterial promoters are well known in the art and are described, *e.g.*, in Sambrook *et al.* and Ausubel *et al.* Bacterial expression systems for expressing the MRP protein are available in, *e.g.*, *E. coli, Bacillus sp.,* and *Salmonella* (Palva *et al.*, *Gene* 22:229-235 (1983); Mosbach *et al.*, *Nature* 302:543-545 (1983). Kits for such expression systems are commercially available. Eukaryotic expression systems for mammalian cells, yeast, and insect cells are well known in the art and are also commercially available. In one embodiment, the eukaryotic expression vector is an adenoviral vector, an adeno-associated vector, or a retroviral vector.

For therapeutic applications, MRP nucleic acids are introduced into a cell, *in vitro*, *in vivo*, or *ex vivo*, using any of a large number of methods including, but not limited to, infection with viral vectors, liposome-based methods, biolistic particle acceleration (the gene gun), and naked DNA injection. Such therapeutically useful nucleic acids include, but are not limited to, coding sequences for full-length MRP, coding sequences for an MRP fragment, domain, derivative, or variant, MRP antisense sequences, and MRP ribozymes. Typically, such sequences will be operably linked to a promoter, but in numerous applications a nucleic acid will be administered to a cell that is itself directly therapeutically effective, *e.g.*, certain antisense or ribozyme molecules.

The promoter used to direct expression of a heterologous nucleic acid depends on the particular application. The promoter is optionally positioned about the same distance from the heterologous transcription start site as it is from the transcription start site in its natural setting. As is known in the art, however, some variation in this distance can be accommodated without loss of promoter function.

In addition to the promoter, the expression vector typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the MRP-encoding nucleic acid in host cells. A typical expression cassette thus contains a promoter operably linked to the nucleic acid sequence encoding an MRP polypeptide, and signals required for efficient polyadenylation of the transcript, ribosome binding sites, and translation termination. The nucleic acid sequence encoding an MRP polypeptide may be linked to a cleavable signal peptide sequence to promote secretion of the encoded protein by the transfected cell. Such signal peptides would

include, among others, the signal peptides from tissue plasminogen activator, insulin, and neuron growth factor, and juvenile hormone esterase of *Heliothis virescens*. Additional elements of the cassette may include enhancers and, if genomic DNA is used as the structural gene, introns with functional splice donor and acceptor sites.

In addition to a promoter sequence, the expression cassette should also contain a transcription termination region downstream of the structural gene to provide for efficient termination. The termination region may be obtained from the same gene as the promoter sequence or may be obtained from different genes.

The particular expression vector used to transport the genetic information into the cell is not particularly critical. Any of the conventional vectors used for expression in eukaryotic or prokaryotic cells may be used. Standard bacterial expression vectors include plasmids such as pBR322 based plasmids, pSKF, pET23D, and fusion expression systems such as GST and LacZ. Epitope tags can also be added to recombinant proteins to provide convenient methods of isolation, *e.g.*, c-myc, HA-tag, 6-His tag, maltose binding protein, VSV-G tag, anti-DYKDDDDK tag, or any such tag, a large number of which are well known to those of skill in the art.

Expression vectors containing regulatory elements from eukaryotic viruses are typically used in eukaryotic expression vectors, *e.g.*, SV40 vectors, papilloma virus vectors, and vectors derived from Epstein-Barr virus. Other exemplary eukaryotic vectors include pMSG, pAV009/A^+, pMTO10/A^+, pMAMneo-5, baculovirus pDSVE, and any other vector allowing expression of proteins under the direction of the CMV promoter, SV40 early promoter, SV40 later promoter, metallothionein promoter, murine mammary tumor virus promoter, Rous sarcoma virus promoter, polyhedrin promoter, or other promoters shown effective for expression in eukaryotic cells.

Some expression systems have markers that provide gene amplification, such as neomycin, thymidine kinase, hygromycin B phosphotransferase, and dihydrofolate reductase. Alternatively, high yield expression systems not involving gene amplification are also suitable, such as using a baculovirus vector in insect cells, with a sequence encoding an MRP polypeptide under the direction of the polyhedrin promoter or other strong baculovirus promoters.

The elements that are typically included in expression vectors also include a replicon that functions in *E. coli*, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of eukaryotic sequences. The

particular antibiotic resistance gene chosen is not critical, any of the many resistance genes known in the art are suitable. The prokaryotic sequences are optionally chosen such that they do not interfere with the replication of the DNA in eukaryotic cells, if necessary.

Standard transfection methods are used to produce bacterial, mammalian, yeast or insect cell lines that express large quantities of an MRP protein, which are then purified using standard techniques (*see, e.g.*, Colley *et al.*, *J. Biol. Chem.* 264:17619-17622 (1989); *Guide to Protein Purification*, in *Methods in Enzymology*, vol. 182 (Deutscher, ed., 1990)). Transformation of eukaryotic and prokaryotic cells are performed according to standard techniques (*see, e.g.*, Morrison, *J. Bact.* 132:349-351 (1977); Clark-Curtiss & Curtiss, *Methods in Enzymology* 101:347-362 (Wu *et al.*, eds, 1983).

Any of the well known procedures for introducing foreign nucleotide sequences into host cells may be used to introduce the expression vector. These include the use of reagents such as Superfect (Qiagen), liposomes, calcium phosphate transfection, polybrene, protoplast fusion, electroporation, microinjection, plasmid vectors, viral vectors, biolistic particle acceleration (the gene gun), or any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (*see, e.g.*, Sambrook *et al., supra*). It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing an MRP gene.

After the expression vector is introduced into the cells, the transfected cells are cultured under conditions favoring expression of the MRP polypeptide, which is recovered from the culture using standard techniques identified below. Methods of culturing prokaryotic or eukaryotic cells are well known and are taught, *e.g.*, in Ausubel *et al.*, Sambrook *et al.*, and in Freshney, Culture of Animal Cells, 3d. Ed., (1993), A Wiley-Liss Publication.

IV. Purification of MRP polypeptides

Either naturally occurring or recombinant MRP polypeptides can be purified for use in functional assays, binding assays, diagnostic assays, and other applications. Naturally occurring MRP polypeptides are purified, *e.g.*, from mammalian tissue such as brain, kidneys, lung, liver or other tissues, or any other source of an MRP

homolog. Recombinant MRP polypeptides are purified from any suitable bacterial or eukaryotic expression system, *e.g.*, CHO cells or insect cells.

MRP proteins may be purified to substantial purity by standard techniques, including selective precipitation with such substances as ammonium sulfate; column chromatography, immunopurification methods, and others (*see, e.g.*, Scopes, *Protein Purification: Principles and Practice* (1982); U.S. Patent No. 4,673,641; Ausubel *et al., supra*; and Sambrook *et al., supra*).

A number of procedures can be employed when recombinant MRP polypeptide is being purified. For example, proteins having established molecular adhesion properties can be reversibly fused to the MRP polypeptide. With the appropriate ligand, an MRP polypeptide can be selectively adsorbed to a purification column and then freed from the column in a relatively pure form. The fused protein is then removed by enzymatic activity. MRP proteins can also be purified using immunoaffinity columns.

A. Purification of MRP Protein from Recombinant Cells

Recombinant proteins are expressed by transformed bacteria or eukaryotic cells such as CHO cells or insect cells in large amounts, typically after promoter induction; but expression can be constitutive. Promoter induction with IPTG is one example of an inducible promoter system. Cells are grown according to standard procedures in the art. Fresh or frozen cells are used for isolation of protein.

Proteins expressed in bacteria may form insoluble aggregates ("inclusion bodies"). Several protocols are suitable for purification of MRP inclusion bodies. For example, purification of inclusion bodies typically involves the extraction, separation and/or purification of inclusion bodies by disruption of bacterial cells, *e.g.*, by incubation in a buffer of 50 mM TRIS/HCL pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mM ATP, and 1 mM PMSF. The cell suspension can be lysed using 2-3 passages through a French Press, homogenized using a Polytron (Brinkman Instruments) or sonicated on ice. Alternate methods of lysing bacteria are apparent to those of skill in the art (*see, e.g.*, Sambrook *et al., supra*; Ausubel *et al., supra*).

If necessary, the inclusion bodies are solubilized, and the lysed cell suspension is typically centrifuged to remove unwanted insoluble matter. Proteins that formed the inclusion bodies may be renatured by dilution or dialysis with a compatible buffer. Suitable solvents include, but are not limited to, urea (from about 4 M to about 8

M), formamide (at least about 80%, volume/volume basis), and guanidine hydrochloride (from about 4 M to about 8 M). Some solvents which are capable of solubilizing aggregate-forming proteins, for example SDS (sodium dodecyl sulfate) and 70% formic acid, are inappropriate for use in this procedure due to the possibility of irreversible denaturation of the proteins, accompanied by a lack of immunogenicity and/or activity. Although guanidine hydrochloride and similar agents are denaturants, this denaturation is not irreversible and renaturation may occur upon removal (by dialysis, for example) or dilution of the denaturant, allowing re-formation of immunologically and/or biologically active protein. Other suitable buffers are known to those skilled in the art. MRP polypeptides are separated from other bacterial proteins by standard separation techniques, *e.g.*, with Ni-NTA agarose resin.

Alternatively, it is possible to purify MRP polypeptides from bacteria periplasm. After lysis of the bacteria, when an MRP protein is exported into the periplasm of the bacteria, the periplasmic fraction of the bacteria can be isolated by cold osmotic shock in addition to other methods known to skill in the art. To isolate recombinant proteins from the periplasm, the bacterial cells are centrifuged to form a pellet. The pellet is resuspended in a buffer containing 20% sucrose. To lyse the cells, the bacteria are centrifuged and the pellet is resuspended in ice-cold 5 mM $MgSO_4$ and kept in an ice bath for approximately 10 minutes. The cell suspension is centrifuged and the supernatant decanted and saved. The recombinant proteins present in the supernatant can be separated from the host proteins by standard separation techniques well known to those of skill in the art.

B. Standard protein separation techniques for purifying MRP polypeptides

1. Solubility fractionation

Often as an initial step, particularly if the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. The preferred salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol includes adding

saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This concentration will precipitate the most hydrophobic of proteins. The precipitate is then discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, either through dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

2. Size differential filtration

The molecular weight of an MRP protein can be used to isolate it from proteins of greater and lesser size using ultrafiltration through membranes of different pore size (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed as described below.

3. Column chromatography

MRP proteins can also be separated from other proteins on the basis of their size, net surface charge, hydrophobicity, and affinity for heterologous molecules. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these methods are well known in the art. It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (*e.g.*, Pharmacia Biotech).

V. Antibodies to MRP Family Members

In numerous embodiments of the present invention, antibodies that specifically bind to MRP polypeptides will be used. Such antibodies have numerous applications, including for the modulation of MRP activity and for immunoassays to detect MRP, and variants, derivatives, fragments, *etc.* of MRP. Immunoassays can be used to qualitatively or quantitatively analyze the MRP polypeptide. A general overview

of the applicable technology can be found in Harlow & Lane, *Antibodies: A Laboratory Manual* (1988).

Methods of producing polyclonal and monoclonal antibodies that react specifically with MRP polypeptides are known to those of skill in the art (*see, e.g.*, Coligan, *Current Protocols in Immunology* (1991); Harlow & Lane, *supra*; Goding, *Monoclonal Antibodies: Principles and Practice* (2d ed. 1986); and Kohler & Milstein, *Nature* 256:495-497 (1975)). Such techniques include antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors, as well as preparation of polyclonal and monoclonal antibodies by immunizing rabbits or mice (*see, e.g.*, Huse *et al.*, *Science* 246:1275-1281 (1989); Ward *et al.*, *Nature* 341:544-546 (1989)).

A number of MRP-comprising immunogens may be used to produce antibodies specifically reactive with an MRP polypeptide. For example, a recombinant MRP protein, or an antigenic fragment thereof, is isolated as described herein. Recombinant protein can be expressed in eukaryotic or prokaryotic cells as described above, and purified as generally described above. Recombinant protein is the preferred immunogen for the production of monoclonal or polyclonal antibodies. Alternatively, a synthetic peptide derived from the sequences disclosed herein and conjugated to a carrier protein can be used an immunogen. Naturally occurring protein may also be used either in pure or impure form. The product is then injected into an animal capable of producing antibodies. Either monoclonal or polyclonal antibodies may be generated, for subsequent use in immunoassays to measure the protein.

Methods of production of polyclonal antibodies are known to those of skill in the art. An inbred strain of mice (*e.g.*, BALB/C mice) or rabbits is immunized with the protein using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the MRP polypeptide. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done if desired (*see,* Harlow & Lane, *supra*).

Monoclonal antibodies may be obtained by various techniques familiar to those skilled in the art. Briefly, spleen cells from an animal immunized with a desired antigen are immortalized, commonly by fusion with a myeloma cell (*see* Kohler &

Milstein, *Eur. J. Immunol.* 6:511-519 (1976)). Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for the antigen, and yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse *et al.*, *Science* 246:1275-1281 (1989).

Monoclonal antibodies and polyclonal sera are collected and titered against the immunogen protein in an immunoassay, for example, a solid phase immunoassay with the immunogen immobilized on a solid support. Typically, polyclonal antisera with a titer of 10^4 or greater are selected and tested for their cross reactivity against non-MRP proteins, or even related proteins from other organisms, using a competitive binding immunoassay. Specific polyclonal antisera and monoclonal antibodies will usually bind with a K_d of at least about 0.1 mM, more usually at least about 1 μM, optionally at least about 0.1 μM or better, and optionally 0.01 μM or better.

Using MRP-specific antibodies, individual MRP proteins can be detected by a variety of immunoassay methods. For a review of immunological and immunoassay procedures, see *Basic and Clinical Immunology* (Stites & Terr eds., 7th ed. 1991). Moreover, the immunoassays of the present invention can be performed in any of several configurations, which are reviewed extensively in *Enzyme Immunoassay* (Maggio, ed., 1980); and Harlow & Lane, *supra*.

A. Immunological Binding Assays

MRP proteins can be detected and/or quantified using any of a number of well recognized immunological binding assays (*see, e.g.,* U.S. Patent Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also *Methods in Cell Biology: Antibodies in Cell Biology*, volume 37 (Asai, ed. 1993); *Basic and Clinical Immunology* (Stites & Terr, eds., 7th ed. 1991). Immunological binding assays (or immunoassays) typically use an antibody that specifically binds to a protein or antigen of choice (in this case an MRP protein or an antigenic subsequence thereof). The antibody (*e.g.*, anti-MRP) may be produced by any of a number of means well known to those of skill in the art and as described above.

Immunoassays also often use a labeling agent to specifically bind to and label the complex formed by the antibody and antigen. The labeling agent may itself be one of the moieties comprising the antibody/antigen complex. Thus, the labeling agent may be a labeled MRP polypeptide or a labeled anti-MRP antibody. Alternatively, the labeling agent may be a third moiety, such a secondary antibody, that specifically binds to the antibody/MRP complex (a secondary antibody is typically specific to antibodies of the species from which the first antibody is derived). Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G, may also be used as the label agent. These proteins exhibit a strong nonimmunogenic reactivity with immunoglobulin constant regions from a variety of species (*see, e.g.,* Kronval *et al., J. Immunol.* 111:1401-1406 (1973); Akerstrom *et al., J. Immunol.* 135:2589-2542 (1985)). The labeling agent can be modified with a detectable moiety, such as biotin, to which another molecule can specifically bind, such as streptavidin. A variety of detectable moieties are well known to those skilled in the art.

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, optionally from about 5 minutes to about 24 hours. However, the incubation time will depend upon the assay format, antigen, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

1. Noncompetitive assay formats

Immunoassays for detecting an MRP protein in a sample may be either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of antigen is directly measured. In one preferred "sandwich" assay, for example, the anti-MRP antibodies can be bound directly to a solid substrate on which they are immobilized. These immobilized antibodies then capture the MRP protein present in the test sample. The MRP protein is thus immobilized is then bound by a labeling agent, such as a second MRP antibody bearing a label. Alternatively, the second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second or third antibody is typically modified with a detectable moiety, such as biotin, to which another molecule specifically binds, *e.g.*, streptavidin, to provide a detectable moiety.

2. Competitive assay formats

In competitive assays, the amount of MRP protein present in the sample is measured indirectly by measuring the amount of a known, added (exogenous) MRP protein displaced (competed away) from an anti-MRP antibody by the unknown MRP protein present in a sample. In one competitive assay, a known amount of MRP protein is added to a sample and the sample is then contacted with an antibody that specifically binds to the MRP protein. The amount of exogenous MRP protein bound to the antibody is inversely proportional to the concentration of MRP protein present in the sample. In a particularly preferred embodiment, the antibody is immobilized on a solid substrate. The amount of MRP protein bound to the antibody may be determined either by measuring the amount of MRP protein present in an MRP/antibody complex, or alternatively by measuring the amount of remaining uncomplexed protein. The amount of MRP protein may be detected by providing a labeled MRP molecule.

A hapten inhibition assay is another preferred competitive assay. In this assay the known MRP protein is immobilized on a solid substrate. A known amount of anti-MRP antibody is added to the sample, and the sample is then contacted with the immobilized MRP. The amount of anti-MRP antibody bound to the known immobilized MRP protein is inversely proportional to the amount of MRP protein present in the sample. Again, the amount of immobilized antibody may be detected by detecting either the immobilized fraction of antibody or the fraction of the antibody that remains in solution. Detection may be direct where the antibody is labeled or indirect by the subsequent addition of a labeled moiety that specifically binds to the antibody as described above.

3. Cross-reactivity determinations

Immunoassays in the competitive binding format can also be used for crossreactivity determinations. For example, a protein at least partially encoded by SEQ ID NO:2, 3, 5 or 7 can be immobilized to a solid support. Proteins (*e.g.*, MRP proteins and homologs) are added to the assay that compete for binding of the antisera to the immobilized antigen. The ability of the added proteins to compete for binding of the antisera to the immobilized protein is compared to the ability of the MRP polypeptide encoded by SEQ ID NO:2, 3, 5 or 7 to compete with itself. The percent cross-reactivity for the above proteins is calculated, using standard calculations. Those antisera with less

than 10% cross-reactivity with each of the added proteins listed above are selected and pooled. The cross-reacting antibodies are optionally removed from the pooled antisera by immunoabsorption with the added considered proteins, *e.g.*, distantly related homologs.

The immunoabsorbed and pooled antisera are then used in a competitive binding immunoassay as described above to compare a second protein, thought to be perhaps an allele or polymorphic variant of an MRP protein, to the immunogen protein (*i.e.*, MRP protein encoded by SEQ ID NO:2, 3, 5 or 7). In order to make this comparison, the two proteins are each assayed at a wide range of concentrations and the amount of each protein required to inhibit 50% of the binding of the antisera to the immobilized protein is determined. If the amount of the second protein required to inhibit 50% of binding is less than 10 times the amount of the protein encoded by SEQ ID NO:2, 3, 5 or 7 that is required to inhibit 50% of binding, then the second protein is said to specifically bind to the polyclonal antibodies generated to an MRP immunogen.

Polyclonal antibodies that specifically bind to an MRP protein from a particular species can be made by subtracting out cross-reactive antibodies using MRP homologs. For example, antibodies specific to human MRP (*e.g.*, antibodies generated against a polypeptide comprising an amino acid sequence of SEQ ID NO: 4 or 6) can be made by subtracting out antibodies that are cross-reactive with mouse MRP (*e.g.*, antibodies that selectively bind to a polypeptide comprising an amino acid sequence of SEQ ID NO:1). In an analogous fashion, antibodies specific to a particular MRP protein can be obtained in organisms with multiple MRP genes.

4. Other assay formats

Western blot (immunoblot) analysis is used to detect and quantify the presence of MRP protein in a sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the antibodies that specifically bind the MRP protein. The anti-MRP polypeptide antibodies specifically bind to the MRP polypeptide on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (*e.g.*, labeled sheep anti-mouse antibodies) that specifically bind to the anti-MRP antibodies.

Other assay formats include liposome immunoassays (LIA), which use liposomes designed to bind specific molecules (*e.g.*, antibodies) and release encapsulated

reagents or markers. The released chemicals are then detected according to standard techniques (*see* Monroe *et al.*, *Amer. Clin. Prod. Rev.* 5:34-41 (1986)).

5. Reduction of nonspecific binding

One of skill in the art will appreciate that it is often desirable to minimize nonspecific binding in immunoassays. Particularly, where the assay involves an antigen or antibody immobilized on a solid substrate it is desirable to minimize the amount of nonspecific binding to the substrate. Means of reducing such nonspecific binding are well known to those of skill in the art. Typically, this technique involves coating the substrate with a proteinaceous composition. In particular, protein compositions such as bovine serum albumin (BSA), nonfat powdered milk, and gelatin are widely used with powdered milk being most preferred.

6. Labels

The particular label or detectable group used in the assay is not a critical aspect of the invention, as long as it does not significantly interfere with the specific binding of the antibody used in the assay. The detectable group can be any material having a detectable physical or chemical property. Such detectable labels have been well-developed in the field of immunoassays and, in general, most any label useful in such methods can be applied to the present invention. Thus, a label is any composition detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Useful labels in the present invention include magnetic beads (*e.g.*, DYNABEADS^{TM}), fluorescent dyes (*e.g.*, fluorescein isothiocyanate, Texas red, rhodamine, and the like), radiolabels (*e.g.*, ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic beads (*e.g.*, polystyrene, polypropylene, latex, *etc.*).

The label may be coupled directly or indirectly to the desired component of the assay according to methods well known in the art. As indicated above, a wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

Nonradioactive labels are often attached by indirect means. Generally, a ligand molecule (*e.g.*, biotin) is covalently bound to the molecule. The ligand then binds

to another molecules (*e.g.*, streptavidin) molecule, which is either inherently detectable or covalently bound to a signal system, such as a detectable enzyme, a fluorescent compound, or a chemiluminescent compound. The ligands and their targets can be used in any suitable combination with antibodies that recognize an MRP protein, or secondary antibodies that recognize anti-MRP.

The molecules can also be conjugated directly to signal generating compounds, *e.g.*, by conjugation with an enzyme or fluorophore. Enzymes of interest as labels will primarily be hydrolases, particularly phosphatases, esterases and glycosidases, or oxidases, particularly peroxidases. Fluorescent compounds include fluorescein and its derivatives, rhodamine and its derivatives, dansyl, umbelliferone, *etc.* Chemiluminescent compounds include luciferin, and 2,3-dihydrophthalazinediones, *e.g.*, luminol. For a review of various labeling or signal producing systems that may be used, *see*, *e.g.*, U.S. Patent No. 4,391,904.

Means of detecting labels are well known to those of skill in the art. Thus, for example, where the label is a radioactive label, means for detection include a scintillation counter or photographic film as in autoradiography. Where the label is a fluorescent label, it may be detected by exciting the fluorochrome with the appropriate wavelength of light and detecting the resulting fluorescence. The fluorescence may be detected visually, by means of photographic film, by the use of electronic detectors such as charge coupled devices (CCDs) or photomultipliers and the like. Similarly, enzymatic labels may be detected by providing the appropriate substrates for the enzyme and detecting the resulting reaction product. Finally simple colorimetric labels may be detected simply by observing the color associated with the label. Thus, in various dipstick assays, conjugated gold often appears pink, while various conjugated beads appear the color of the bead.

Some assay formats do not require the use of labeled components. For instance, agglutination assays can be used to detect the presence of the target antibodies. In this case, antigen-coated particles are agglutinated by samples comprising the target antibodies. In this format, none of the components need be labeled and the presence of the target antibody is detected by simple visual inspection.

VI. Modulating MRP Activity in Cells

A. Assays for Modulators of MRP Proteins

In numerous embodiments of this invention, the level of MRP activity will be modulated in a cell by administering to the cell, *in vivo* or *in vitro*, any of a large number of potential MRP-modulating molecules, *e.g.*, polypeptides, antibodies, amino acids, nucleotides, lipids, carbohydrates, or any organic or inorganic molecule. Such MRP modulators are particularly useful in the prevention or treatment of any of a large number of estrogen-associated diseases or conditions.

To identify molecules capable of modulating MRP, assays will be performed to detect the effect of various compounds on MRP activity. Such assays can involve the identification of compounds that interact with MRP proteins, either physically or genetically, and can thus rely on any of a number of standard methods to detect physical or genetic interactions between compounds. Such assays can also involve the identification of compounds that affect MRP expression, activity or other properties, such as its phosphorylation state or ability to bind other proteins. Such assays can also involve the detection of MRP activity in a cell, either *in vitro* or *in vivo*. Such cell-based assays can be performed in any type of cell, *e.g.*, a cell that naturally expresses MRP, or a cultured cell that produces MRP due to recombinant expression.

B. Assays for MRP-Interacting Compounds

In certain embodiments, assays will be performed to identify molecules that physically or genetically interact with MRP proteins. Such molecules can be any type of molecule, including polypeptides, polynucleotides, amino acids, nucleotides, carbohydrates, lipids, or any other organic or inorganic molecule. Such molecules may represent molecules that normally interact with MRP to effect estrogen signaling, or may be synthetic or other molecules that are capable of interacting with MRP and that can potentially be used to modulate MRP activity in cells, or used as lead compounds to identify classes of molecules that can interact with and/or modulate MRP. Such assays may represent physical binding assays, such as affinity chromatography, immunoprecipitation, two-hybrid screens, or other binding assays, or may represent genetic assays as described *infra*.

In any of the binding or functional assays described herein, *in vivo* or *in vitro*, any MRP protein, or any derivative, variation, homolog, or fragment of an MRP protein, can be used. Preferably, the MRP protein is at least about 70% similar or

identical to SEQ ID NO:1, 4 or 6, or comprises an MRP signature sequence, *e.g.*, as shown in SEQ ID NO:8, 9, 10, or 11. In numerous embodiments, a fragment of an MRP protein is used. For example, a fragment that contains only a myosin head domain, a tail domain, an SH3 domain, an IQ motif, or an MyTH4 domain is used. Such fragments can be used alone, in combination with other MRP fragments, or in combination with sequences from heterologous proteins, *e.g.*, the fragments can be fused to a heterologous polypeptide, thereby forming a chimeric polypeptide.

1. Assays for physical interactions

Compounds that interact with MRP proteins can be isolated based on an ability to specifically bind to an MRP protein or fragment thereof. In numerous embodiments, the MRP protein or protein fragment will be attached to a solid support. In one embodiment, affinity columns are made using the MRP polypeptide, and physically-interacting molecules are identified. It will be apparent to one of skill that chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (*e.g.*, Pharmacia Biotech). In addition, molecules that interact with MRP proteins *in vivo* can be identified by co-immunoprecipitation or other methods, *i.e.*, immunoprecipitating MRP proteins using anti-MRP antibodies from a cell or cell extract, and identifying compounds, *e.g.*, proteins, that are precipitated along with the MRP protein. Such methods are well known to those of skill in the art and are taught, *e.g.*, in Ausubel *et al.*, Sambrook *et al.*, Harlow & Lane, all *supra*.

Two-hybrid screens can also be used to identify polypeptides that interact *in vivo* with an MRP polypeptide or a fragment thereof (Fields *et al.*, *Nature* 340:245-246 (1989)). Such screens comprise two discrete, modular domains of a transcription factor protein, *e.g.*, a DNA binding domain and a transcriptional activation domain, which are produced in a cell as two separate polypeptides, each of which also comprises one of two potentially binding polypeptides. If the two potentially binding polypeptides do in fact interact *in vivo*, then the DNA binding and the transcriptional activating domain of the transcription factor are united, thereby producing expression of a target gene in the cell. The target gene typically encodes an easily detectable gene product, *e.g.*, β-galactosidase, GFP, or luciferase, which can be detected using standard methods. In the present invention, an MRP polypeptide, or fragment thereof, is fused to one of the two domains of the transcription factor, and the potential MRP-binding polypeptides (*e.g.*, encoded by

a cDNA library) are fused to the other domain. Such methods are well known to those of skill in the art, and are taught, *e.g.*, in Ausubel *et al., supra*.

C. Assays for MRP Protein Activity

MRP genes and their alleles and polymorphic variants encode mediators of estrogen receptor signaling, in particular specifically through the estrogen receptor α or, alternatively, specifically through the estrogen receptor β. Accordingly, the activity of MRP polypeptides can be assessed using a variety of *in vitro* and *in vivo* assays to determine functional, chemical, and physical effects, *e.g.*, measuring the binding of MRP to heterologous proteins or molecules, including membrane associated proteins or to heterologous SH3-containing proteins, measuring MRP protein and/or RNA levels, or measuring other aspects of MRP polypeptides, *e.g.*, phosphorylation levels, and the like. Such assays can be used to test for both activators and inhibitors of MRP proteins. Modulators can also be genetically altered versions of MRP proteins, *e.g.*, dominant negative forms of MRP or of proteins that interact with MRP. Such modulators of activity are useful for many diagnostic and therapeutic applications.

The MRP protein of the assay will typically be a recombinant or naturally occurring polypeptide with a sequence of SEQ ID NO:1, 4 or 6, or conservatively modified variants thereof. Alternatively, the MRP protein of the assay will be derived from a eukaryote and include an amino acid subsequence having amino acid sequence identity or similarity to SEQ ID NO:1, 4 or 6. Generally, the amino acid sequence identity will be at least 60%, optionally at least 70% to 85%, optionally at least 90-95%. Optionally, the polypeptide of the assays will comprise a domain of an MRP protein, such as a myosin head domain, an SH3 domain, an IQ domain, a talin domain, a tail domain, or an MyTH4 domain. In certain embodiments, a domain of an MRP protein is bound to a solid substrate and used, *e.g.*, to isolate any molecules that can bind to and/or modulate their activity. In certain embodiments, a domain of an MRP polypeptide is fused to a heterologous polypeptide, thereby forming a chimeric polypeptide. Such chimeric polypeptides are also useful, *e.g.*, in assays to identify modulators of MRP.

Samples or assays that are treated with a potential MRP protein inhibitor or activator are compared to control samples without the test compound, to examine the extent of modulation. Control samples (untreated with activators or inhibitors) are assigned a relative MRP activity value of 100. Inhibition of an MRP protein is achieved when the MRP activity value relative to the control is about 90%, optionally 50%,

optionally 25-0%. Activation of an mMRP protein is achieved when the mMRP activity value relative to the control is 110%, optionally 150%, 200-500%, or 1000-2000%.

The effects of the test compounds upon the function of the polypeptides can be measured by examining any of the parameters described above. Any suitable physiological change that is affected by MRP activity can be used to assess the influence of a test compound on the polypeptides of this invention. When the functional consequences are determined using intact cells or animals, one can also measure a variety of effects that are indicative of estrogen signaling, *e.g.*, inhibition of atherosclerosis or other cardiovascular conditions, inhibition of osteroporosis, *etc.* In numerous embodiments, cells or animals used in such embodiments are estrogen receptor α-specific, *i.e.* they contain mutations in genes encoding estrogen receptor β or other estrogen receptors. In such embodiments, cells or animals that are mutant for an estrogen receptor α-encoding gene can be used as a negative control. In numerous other embodiments, cells or animals used in such embodiments are estrogen receptor β-specific, *i.e.* they contain mutations in genes encoding estrogen receptor α or other estrogen receptors. In such embodiments, cells or animals that are mutant for an estrogen receptor β-encoding gene can be used as a negative control.

D. Identifying estrogen receptor agonists and antagonists

In a preferred embodiment, transcription levels of MRP are measured to assess the effects of a test compound on estrogen signaling, *e.g.*, mediated by an estrogen receptor α or, alternatively, by an estrogen receptor β. A host cell containing an estrogen receptor of interest is contacted with a test compound for a sufficient time to effect any interactions, and the level of MRP gene expression is measured. The amount of time to effect such interactions may be empirically determined, such as by running a time course and measuring the level of transcription as a function of time.

Such assays can be performed using any suitable eukaryotic cells, including, *e.g.*, mammalian cells, insect cells, plant or yeast cells using standard methods. A cell type will be selected that naturally expresses an estrogen receptor, *e.g.*, an estrogen receptor α or, alternatively, by an estrogen receptor β, or which is induced to express an estrogen receptor by, for example, introducing a heterologous polynucleotide encoding the receptor, operably linked to a promoter, into the cell.

The amount of transcription may be measured using any method known to those of skill in the art to be suitable. For example, mRNA expression of the protein of

interest may be detected using, *e.g.*, Northern blots, reverse transcriptase-polymerase chain reaction (RT-PCR), or any standard method, using, *e.g.*, probes or primers designed from SEQ ID NO: 2, 3, 5 or 7. Alternatively, the expression of the MRP can be detected by detecting the level of the polypeptide products using immunoassays or other assays to detect MRP activity. Such assays can use natural forms of MRP or can use an MRP fusion, *e.g.*, MRP transcript fused to a reporter sequence, *e.g.*, a sequence encoding chloramphenicol acetyltransferase, luciferase, β-galactosidase, GFP, or alkaline phosphatase. Alternatively, one of these marker sequences can be operably linked to an MRP promoter. (*See, e.g.*, Mistili & Spector, *Nature Biotechnology* 15:961-964 (1997)).

The amount of transcription is then compared to the amount of transcription in either the same cell in the absence of the test compound, or it may be compared with the amount of transcription in a substantially identical cell that lacks the protein of interest. A substantially identical cell may be derived from the same cells from which the recombinant cell was prepared but which had not been modified by introduction of heterologous DNA. Any difference in the amount of transcription indicates that the test compound has in some manner altered the activity of the protein of interest.

A compound that causes an increase in the amount of MRP expression, as detected by any of the herein-described methods, is a candidate for an estrogen receptor agonist that would be useful, *e.g.*, in the inhibition of osteoporosis, atherosclerosis, Alzheimer's Disease, or Parkinson's Disease, and a compound that causes a decrease in the amount of MRP expression is a candidate for an estrogen-receptor antagonist, which would be useful in the inhibition of, *e.g.*, breast cancer. Candidate agonists or antagonists can be further characterized by any of a number of methods, including, *e.g.*, directly examining their interactions with estrogen receptors, examining their ability to alter MRP expression in cells that do not express estrogen receptors, examining the estrogen receptor specificity of the candidate (*i.e.*, estrogen receptor α vs. β), etc.

In preferred embodiments, the detection of the level of MRP expression in the presence or absence of the test compound is tested simultaneously for a large number of test compounds, *e.g.*, using high throughput screening.

E. Modulators and Binding Compounds

The compounds tested as modulators of an MRP protein can be any small chemical compound, *e.g.*, organic or inorganic compounds, or any biological entity, such

as a protein, sugar, nucleic acid or lipid. Alternatively, modulators can be genetically altered versions of an MRP gene. Typically, test compounds will be small chemical molecules and peptides. Essentially any chemical compound can be used as a potential modulator or binding compound in the assays of the invention, although most often compounds can be dissolved in aqueous or organic (especially DMSO-based) solutions are used. The assays are designed to screen large chemical libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (*e.g.*, in microtiter formats on microtiter plates in robotic assays). It will be appreciated that there are many suppliers of chemical compounds, including Sigma (St. Louis, MO), Aldrich (St. Louis, MO), Sigma-Aldrich (St. Louis, MO), Fluka Chemika-Biochemica Analytika (Buchs, Switzerland) and the like.

In one preferred embodiment, high throughput screening methods involve providing a combinatorial chemical or peptide library containing a large number of potential therapeutic compounds (potential modulator or binding compounds). Such "combinatorial chemical libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as conventional "lead compounds" or can themselves be used as potential or actual therapeutics.

A combinatorial chemical library is a collection of diverse chemical compounds generated by either chemical synthesis or biological synthesis, by combining a number of chemical "building blocks" such as reagents. For example, a linear combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (amino acids) in every possible way for a given compound length (*i.e.*, the number of amino acids in a polypeptide compound). Millions of chemical compounds can be synthesized through such combinatorial mixing of chemical building blocks.

Preparation and screening of combinatorial chemical libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (*see, e.g.,* U.S. Patent No. 5,010,175, Furka, *Int. J. Pept. Prot. Res.* 37:487-493 (1991) and Houghton *et al.*, *Nature* 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used. Such chemistries include, but are not limited to: peptoids (*e.g.*, PCT Publication No. WO 91/19735), encoded peptides (*e.g.*, PCT Publication No. WO 93/20242), random bio-oligomers (*e.g.*,

PCT Publication No. WO 92/00091), benzodiazepines (*e.g.*, U.S. Pat. No. 5,288,514), diversomers such as hydantoins, benzodiazepines and dipeptides (Hobbs *et al.*, *Proc. Natl. Acad. Sci. USA* 90:6909-6913 (1993)), vinylogous polypeptides (Hagihara *et al.*, *J. Amer. Chem. Soc.* 114:6568 (1992)), nonpeptidal peptidomimetics with glucose scaffolding (Hirschmann *et al.*, *J. Amer. Chem. Soc.* 114:9217-9218 (1992)), analogous organic syntheses of small compound libraries (Chen *et al.*, *J. Amer. Chem. Soc.* 116:2661 (1994)), oligocarbamates (Cho *et al.*, *Science* 261:1303 (1993)), and/or peptidyl phosphonates (Campbell *et al.*, *J. Org. Chem.* 59:658 (1994)), nucleic acid libraries (*see,* Ausubel, Berger and Sambrook, all *supra*), peptide nucleic acid libraries (*see, e.g.,* U.S. Patent No. 5,539,083), antibody libraries (*see, e.g.,* Vaughn *et al.*, *Nature Biotechnology*, 14(3):309-314 (1996) and PCT/US96/10287), carbohydrate libraries (*see, e.g.,* Liang *et al.*, *Science*, 274:1520-1522 (1996) and U.S. Patent No. 5,593,853), small organic molecule libraries (*see, e.g.,* benzodiazepines, Baum C&EN, Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent No. 5,549,974; pyrrolidines, U.S. Patent Nos. 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent No. 5,506,337; benzodiazepines, U.S. Patent No. 5,288,514, and the like).

Devices for the preparation of combinatorial libraries are commercially available (*see, e.g.,* 357 MPS, 390 MPS, Advanced Chem Tech, Louisville KY, Symphony, Rainin, Woburn, MA, 433A Applied Biosystems, Foster City, CA, 9050 Plus, Millipore, Bedford, MA). In addition, numerous combinatorial libraries are themselves commercially available (*see, e.g.,* ComGenex, Princeton, N.J., Tripos, Inc., St. Louis, MO, 3D Pharmaceuticals, Exton, PA, Martek Biosciences, Columbia, MD, *etc.*).

1. Solid state and soluble high throughput assays

In one embodiment, the invention provides soluble assays using molecules such as an N-terminal or C-terminal domain either alone or covalently linked to a heterologous protein to create a chimeric molecule. In another embodiment, the invention provides solid phase based *in vitro* assays in a high throughput format, where a domain, chimeric molecule, MRP protein, or cell or tissue expressing an MRP protein is attached to a solid phase substrate.

In the high throughput assays of the invention, it is possible to screen up to several thousand different modulators in a single day. In particular, each well of a microtiter plate can be used to run a separate assay against a selected potential modulator,

or, if concentration or incubation time effects are to be observed, every 5-10 wells can test a single modulator. Thus, a single standard microtiter plate can assay about 100 (*e.g.*, 96) modulators. If 1536 well plates are used, then a single plate can easily assay from about 100 to about 1500 different compounds. It is possible to assay several different plates per day; assay screens for up to about 6,000-20,000 different compounds is possible using the integrated systems of the invention. More recently, microfluidic approaches to reagent manipulation have been developed.

The molecule of interest can be bound to the solid state component, directly or indirectly, via covalent or non covalent linkage, *e.g.*, via a tag. The tag can be any of a variety of components. In general, a molecule which binds the tag (a tag binder) is fixed to a solid support, and the tagged molecule of interest is attached to the solid support by interaction of the tag and the tag binder.

A number of tags and tag binders can be used, based upon known molecular interactions well described in the literature. For example, where a tag has a natural binder, for example, biotin, protein A, or protein G, it can be used in conjunction with appropriate tag binders (avidin, streptavidin, neutravidin, the Fc region of an immunoglobulin, *etc.*) Antibodies to molecules with natural binders such as biotin are also widely available and appropriate tag binders; *see,* SIGMA Immunochemicals 1998 catalogue SIGMA, St. Louis MO).

Similarly, any haptenic or antigenic compound can be used in combination with an appropriate antibody to form a tag/tag binder pair. Thousands of specific antibodies are commercially available and many additional antibodies are described in the literature. For example, in one common configuration, the tag is a first antibody and the tag binder is a second antibody which recognizes the first antibody.

Synthetic polymers, such as polyurethanes, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, and polyacetates can also form an appropriate tag or tag binder. Many other tag/tag binder pairs are also useful in assay systems described herein, as would be apparent to one of skill upon review of this disclosure.

Common linkers such as peptides, polyethers, and the like can also serve as tags, and include polypeptide sequences, such as poly-gly sequences of between about 5 and 200 amino acids. Such flexible linkers are known to persons of skill in the art. For example, poly(ethelyne glycol) linkers are available from Shearwater Polymers, Inc.

Huntsville, Alabama. These linkers optionally have amide linkages, sulfhydryl linkages, or heterofunctional linkages.

Tag binders are fixed to solid substrates using any of a variety of methods currently available. Solid substrates are commonly derivatized or functionalized by exposing all or a portion of the substrate to a chemical reagent which fixes a chemical group to the surface which is reactive with a portion of the tag binder. For example, groups which are suitable for attachment to a longer chain portion would include amines, hydroxyl, thiol, and carboxyl groups. Aminoalkylsilanes and hydroxyalkylsilanes can be used to functionalize a variety of surfaces, such as glass surfaces. The construction of such solid phase biopolymer arrays is well described in the literature. *See, e.g.*, Merrifield, *J. Am. Chem. Soc.* 85:2149-2154 (1963) (describing solid phase synthesis of, *e.g.*, peptides); Geysen *et al., J. Immun. Meth.* 102:259-274 (1987) (describing synthesis of solid phase components on pins); Frank & Doring, *Tetrahedron* 44:60316040 (1988) (describing synthesis of various peptide sequences on cellulose disks); Fodor *et al., Science*, 251:767-777 (1991); Sheldon *et al., Clinical Chemistry* 39(4):718-719 (1993); and Kozal *et al., Nature Medicine* 2(7):753759 (1996) (all describing arrays of biopolymers fixed to solid substrates). Nonchemical approaches for fixing tag binders to substrates include other common methods, such as heat, cross-linking by UV radiation, and the like.

2. Computer-based assays

Yet another assay for compounds that modulate MRP protein activity involves computer assisted drug design, in which a computer system is used to generate a three-dimensional structure of an MRP protein based on the structural information encoded by its amino acid sequence. The input amino acid sequence interacts directly and actively with a pre-established algorithm in a computer program to yield secondary, tertiary, and quaternary structural models of the protein. The models of the protein structure are then examined to identify regions of the structure that have the ability to bind. These regions are then used to identify compounds that bind to the protein.

The three-dimensional structural model of the protein is generated by entering protein amino acid sequences of at least 10 amino acid residues or corresponding nucleic acid sequences encoding an MRP polypeptide into the computer system. The nucleotide sequence encoding the polypeptide preferably comprises SEQ ID NO:2, 3, 5, 7 or conservatively modified versions thereof. The amino acid sequence, preferably

comprising SEQ ID NO:1, 4, 6, or conservatively modifies versions thereof, represents the primary sequence or subsequence of the protein, which encodes the structural information of the protein. At least 10 residues of the amino acid sequence (or a nucleotide sequence encoding 10 amino acids) are entered into the computer system from computer keyboards, computer readable substrates that include, but are not limited to, electronic storage media (*e.g.*, magnetic diskettes, tapes, cartridges, and chips), optical media (*e.g.*, CD ROM), information distributed by Internet sites, and by RAM. The three-dimensional structural model of the protein is then generated by the interaction of the amino acid sequence and the computer system, using software known to those of skill in the art.

The amino acid sequence represents a primary structure that encodes the information necessary to form the secondary, tertiary and quaternary structure of the protein of interest. The software looks at certain parameters encoded by the primary sequence to generate the structural model. These parameters are referred to as "energy terms," and primarily include electrostatic potentials, hydrophobic potentials, solvent accessible surfaces, and hydrogen bonding. Secondary energy terms include van der Waals potentials. Biological molecules form the structures that minimize the energy terms in a cumulative fashion. The computer program is therefore using these terms encoded by the primary structure or amino acid sequence to create the secondary structural model.

The tertiary structure of the protein encoded by the secondary structure is then formed on the basis of the energy terms of the secondary structure. The user at this point can enter additional variables such as whether the protein is membrane bound or soluble, its location in the body, and its cellular location, *e.g.*, cytoplasmic, surface, or nuclear. These variables along with the energy terms of the secondary structure are used to form the model of the tertiary structure. In modeling the tertiary structure, the computer program matches hydrophobic faces of secondary structure with like, and hydrophilic faces of secondary structure with like.

Once the structure has been generated, potential modulator binding regions are identified by the computer system. Three-dimensional structures for potential modulators are generated by entering amino acid or nucleotide sequences or chemical formulas of compounds, as described above. The three-dimensional structure of the potential modulator is then compared to that of the MRP protein to identify compounds that bind to the protein. Binding affinity between the protein and compound is

determined using energy terms to determine which compounds have an enhanced probability of binding to the protein.

Computer systems are also used to screen for mutations, polymorphic variants, alleles and interspecies homologs of MRP genes. Such mutations can be associated with disease states or genetic traits. As described above, GeneChip™ and related technology can also be used to screen for mutations, polymorphic variants, alleles and interspecies homologs. Once the variants are identified, diagnostic assays can be used to identify patients having such mutated genes. Identification of the mutated MRP genes involves receiving input of a first nucleic acid sequence of SEQ ID NO:2, 3, 5, 7, or a first amino acid sequence of SEQ ID NO:1, 4, 6, and conservatively modified versions thereof. The sequence is entered into the computer system as described above. The first nucleic acid or amino acid sequence is then compared to a second nucleic acid or amino acid sequence that has substantial identity to the first sequence. The second sequence is entered into the computer system in the manner described above. Once the first and second sequences are compared, nucleotide or amino acid differences between the sequences are identified. Such sequences can represent allelic differences in various MRP genes, and mutations associated with disease states and genetic traits.

VII. Modulating MRP activity/expression to treat diseases or conditions.

In numerous embodiments of this invention, a compound, *e.g.*, nucleic acid, polypeptide, or other molecule is administered to a patient, *in vivo* or *ex vivo*, to effect a change in MRP activity or expression in the patient. Such compounds can be nucleic acids encoding full length MRP polypeptides, *e.g.*, as shown as SEQ ID NO:1, or any derivative, fragment, or variant thereof, operably linked to a promoter. Suitable nucleic acids also include inhibitory sequences such as antisense or ribozyme sequences, which can be delivered in, *e.g.*, an expression vector operably linked to a promoter, or can be delivered directly. Also, any nucleic acid that encodes a polypeptide that modulates the expression of MRP can be used. In general, nucleic acids can be delivered to cells using any of a large number of vectors or methods, *e.g.*, retroviral, adenoviral, or adeno-associated virus vectors, liposomal formulations, naked DNA injection, and others. All of these methods are well known to those of skill in the art.

Proteins can also be delivered to a patient to modulate MRP activity. In preferred embodiments, a polyclonal or monoclonal antibody that specifically binds to MRP, particularly to a myosin head domain, a tail domain, an SH3 domain, a MyTH4

domain, an IQ motif, or a talin domain, of an MRP polypeptide, is delivered. In addition, any polypeptide that interacts with and/or modulates MRP activity can be used, *e.g.*, a polypeptide that is identified using the presently described assays, or any dominant negative form of MRP or an MRP-interacting protein. In addition, polypeptides that affect MRP expression, as identified using the methods described herein, can be used.

Further, any compound that is found to or designed to interact with and/or modulate the activity of MRP can be used. For example, any compound that is found, using the methods described herein, to bind to or modulate the activity of MRP can be used.

Any of the above-described molecules can be used to increase or decrease the expression or activity of MRP, or to otherwise affect the properties and/or behavior of MRP polypeptides or polynucleotides, *e.g.*, stability, phosphorylation, protein activity, interactions with other proteins, *etc.* The present compounds can thus be used to treat any of a number of diseases, including, but not limited to osteoporosis, atherosclerosis and other cardiovascular diseases, Alzheimer's Disease, Parkinson's Disease, and breast and other cancers.

A. Administration and Pharmaceutical Compositions

Administration of any of the present molecules can be achieved by any of the routes normally used for introducing a modulator compound into ultimate contact with the tissue to be treated. The modulators are administered in any suitable manner, optionally with pharmaceutically acceptable carriers. Suitable methods of administering such modulators are available and well known to those of skill in the art, and, although more than one route can be used to administer a particular composition, a particular route can often provide a more immediate and more effective reaction than another route.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. Accordingly, there is a wide variety of suitable formulations of pharmaceutical compositions of the present invention (*see, e.g., Remington's Pharmaceutical Sciences*, 17th ed. 1985)).

The MRP modulators, alone or in combination with other suitable components, can be made into aerosol formulations (*i.e.*, they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like.

Formulations suitable for administration include aqueous and nonaqueous solutions, isotonic sterile solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic, and aqueous and nonaqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. In the practice of this invention, compositions can be administered, for example, orally, nasally, topically, intravenously, intraperitoneally, intravesically or intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials. Solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described. The modulators can also be administered as part a of prepared food or drug.

The dose administered to a patient, in the context of the present invention should be sufficient to effect a beneficial response in the subject over time. The dose will be determined by the efficacy of the particular modulators employed and the condition of the subject, as well as the body weight or surface area of the area to be treated. The size of the dose also will be determined by the existence, nature, and extent of any adverse side-effects that accompany the administration of a particular compound or vector in a particular subject.

In determining the effective amount of the modulator to be administered, a physician may evaluate circulating plasma levels of the modulator, modulator toxicities, and the production of anti-modulator antibodies. In general, the dose equivalent of a modulator is from about 1 ng/kg to 10 mg/kg for a typical subject.

For administration, modulators of the present invention can be administered at a rate determined by the LD-50 of the modulator, and the side-effects of the compound at various concentrations, as applied to the mass and overall health of the subject. Administration can be accomplished via single or divided doses.

VIII. ***Transgenic Animals***

Transgenic and chimeric non-human mammals and methods for generating them are described below. The mammals are useful, *inter alia*, for testing the function of MRP *in vivo*, to generate models for the study of estrogen-associated diseases and conditions, and for the development of potential treatments for MRP related diseases and conditions.

Transgenic and chimeric non-human mammals are generated that contain cells that lack at least one functional endogenous allele for MRP. A "chimeric animal" includes some cells that lack the functional MRP gene of interest and other cells that do not have the inactivated gene. A "transgenic animal," in contrast, is made up of cells that have all incorporated the specific modification which renders the MRP gene inactive or otherwise altered. While a transgenic animal is typically always capable of transmitting the mutant MRP gene to its progeny, the ability of a chimeric animal to transmit the mutation depends upon whether the inactivated gene is present in the animal's germ cells. The modifications that inactivate or otherwise alter the MRP gene can include, for example, insertions, deletions, or substitutions of one or more nucleotides. The modifications can interfere with transcription of the gene itself, with translation and/or stability of the resulting mRNA, or can cause the gene to encode an inactive or otherwise altered MRP polypeptide, *e.g.*, an MRP polypeptide with modified binding properties.

The claimed methods are useful for producing transgenic and chimeric animals of most vertebrate species. Such species include, but are not limited to, nonhuman mammals, including rodents such as mice and rats, rabbits, ovines such as sheep and goats, porcines such as pigs, and bovines such as cattle and buffalo. Methods of obtaining transgenic animals are described in, for example, Puhler, A., Ed., *Genetic Engineering of Animals*, VCH Publ., 1993; Murphy and Carter, Eds., *Transgenesis Techniques : Principles and Protocols* (*Methods in Molecular Biology*, Vol. 18), 1993; and Pinkert, CA, Ed., *Transgenic Animal Technology : A Laboratory Handbook*, Academic Press, 1994.

In preferred embodiments, transgenic mice will be produced as described in Thomas *et al.* (1999) *Immunol.* 163:978-84; Kanakaraj *et al.* (1998) *J. Exp. Med.* 187:2073-9; or Yeh *et al.*, (1997) *Immunity* 7:715-725.

Typically, a modified mMRP gene is introduced, *e.g.*, by homologous recombination, into embryonic stem cells (ES), which are obtained from preimplantation embryos and cultured *in vitro*. *See, e.g.,* Hooper, ML, *Embryonal Stem Cells: Introducing Planned Changes into the Animal Germline* (Modern Genetics, v. 1), Int'l. Pub. Distrib., Inc., 1993; Bradley *et al.* (1984) *Nature* 309:255-258. Subsequently, the transformed ES cell is combined with a blastocyst from a non-human animal, *e.g.*, a mouse. The ES cells colonize the embryo and in some embryos form the germ line of the resulting chimeric animal. *See*, Jaenisch (1988) *Science* 240: 1468-1474. Alternatively, ES cells or somatic cells that can reconstitute an organism ("somatic repopulating cells")

can be used as a source of nuclei for transplantation into an enucleated fertilized oocyte giving rise to a transgenic mammal. *See, e.g.*, Wilmut *et al.* (1997) *Nature* 385:810-813.

Other methods for obtaining a transgenic or chimeric animal having a mutant MRP gene in its genome is to contact fertilized oocytes with a vector that includes a polynucleotide that encodes a modified, *e.g.*, inactive, MRP polypeptide. In some animals, such as mice, fertilization is typically performed *in vivo* and fertilized ova are surgically removed. In other animals, particularly bovines, it is preferably to remove ova from live or slaughterhouse animals and fertilize the ova *in vitro*. *See* DeBoer *et al.*, WO 91/08216. *In vitro* fertilization permits the modifications to be introduced into substantially synchronous cells.

Fertilized oocytes are typically cultured *in vitro* until a pre-implantation embryo is obtained containing about 16-150 cells. The 16-32 cell stage of an embryo is described as a morula, whereas pre-implantation embryos containing more than 32 cells are termed blastocysts. These embryos show the development of a blastocoel cavity, typically at the 64 cell stage. The presence of a desired MRP mutation in the cells of the embryo can be detected by methods known to those of skill in the art, *e.g.*, Southern blotting, PCR, DNA sequencing, or other standard methods. Methods for culturing fertilized oocytes to the pre-implantation stage are described, *e.g.*, by Gordon *et al.* (1984) *Methods Enzymol.* 101: 414; Hogan *et al. Manipulation of the Mouse Embryo: A Laboratory Manual*, C.S.H.L. N.Y. (1986) (mouse embryo); Hammer *et al.* (1985) *Nature* 315: 680 (rabbit and porcine embryos); Gandolfi *et al.* (1987) *J. Reprod. Fert.* 81:23-28; Rexroad *et al.* (1988) *J. Anim. Sci.* 66:947-953 (ovine embryos) and Eyestone *et al.* (1989) *J. Reprod. Fert.* 85:715-720; Camous *et al.* (1984) *J. Reprod. Fert.* 72:779-785; and Heyman *et al.* (1987) *Theriogenology* 27:5968 (bovine embryos). Pre-implantation embryos may also be stored frozen for a period pending implantation.

Pre-implantation embryos are transferred to an appropriate female resulting in the birth of a transgenic or chimeric animal, depending upon the stage of development when the transgene is integrated. Chimeric mammals can be bred to form true germline transgenic animals. Chimeric mice and germline transgenic mice can also be ordered from commercial sources (*e.g.*, Deltagen, San Carlos, CA).

Other methods for introducing mutations into mammalian cells or animals include recombinase systems, which can be employed to delete all or a portion of a locus of interest. Examples of recombinase systems include, the cre/lox system of bacteriophage P1 (*see, e.g.*, Gu *et al.* (1994) *Science* 265:103-106; Terry *et al.* (1997)

Transgenic Res. 6:349-356) and the FLP/FRT site specific integration system (*see, e.g.*, Dymecki (1996) *Proc. Natl. Acad. Sci. USA* 93:6191-6196). In these systems, sites recognized by the particular recombinase are typically introduced into the genome at a position flanking the portion of the gene that is to be deleted. Introduction of the recombinase into the cells then catalyzes recombination which deletes from the genome the polynucleotide sequence that is flanked by the recombination sites. If desired, one can obtain animals in which only certain cell types lack the MRP gene of interest, *e.g.*, by using a tissue specific promoter to drive the expression of the recombinase. *See, e.g.*, Tsien *et al.* (1996) *Cell* 87:1317-26; Brocard *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:10887-10890; Wang *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:3932-6; Meyers *et al.* (1998) *Nat. Genet.* 18:136-41).

The presence of any mutation in an MRP gene in a cell or animal can be detected using any method described herein, *e.g.*, Southern blot, PCR, or DNA sequencing. *See, e.g.*, Ausubel *et al.*, *supra*.

IX. Kits

MRP genes and their homologs are useful tools for a number of applications, including, but not limited to, identifying estrogen-responsive cells, in particular cells responding to estrogen via an estrogen receptor α or, alternatively, via an estrogen receptor β, for forensics and paternity determinations, and for treating any of a large number of estrogen-associated diseases, such as osteoporosis, atherosclerosis, Alzheimer's Disease, Parkinson's Disease, or breast or other cancers. MRP specific reagents that specifically hybridize to MRP nucleic acids, such as MRP probes and primers, and MRP specific reagents that specifically bind to or modulate the activity of an MRP protein, *e.g.*, MRP antibodies or other compounds can thus be provided in a kit for the practice of any of the applications described herein.

Nucleic acid assays for the presence of DNA and RNA for an MRP polynucleotide in a sample include numerous techniques known to those skilled in the art, such as Southern analysis, Northern analysis, dot blots, RNase protection, S1 analysis, amplification techniques such as PCR and LCR, and *in situ* hybridization. In *in situ* hybridization, for example, the target nucleic acid is liberated from its cellular surroundings in such as to be available for hybridization within the cell while preserving the cellular morphology for subsequent interpretation and analysis. The following articles provide an overview of the art of *in situ* hybridization: Singer *et al.*, *Biotechniques* 4:230-

250 (1986); Haase *et al., Methods in Virology*, vol. VII, pp. 189-226 (1984); and *Nucleic Acid Hybridization: A Practical Approach* (Hames *et al.*, eds. 1987). In addition, an MRP protein can be detected with the various immunoassay techniques described above. The test sample is typically compared to both a positive control (*e.g.*, a sample expressing a recombinant MRP protein) and a negative control.

The present invention also provides kits for screening for modulators of MRP proteins or nucleic acids. Such kits can be prepared from readily available materials and reagents. For example, such kits can comprise any one or more of the following materials: MRP nucleic acids or proteins, reaction tubes, and instructions for testing MRP activity. Optionally, the kit contains a biologically active MRP protein. A wide variety of kits and components can be prepared according to the present invention, depending upon the intended user of the kit and the particular needs of the user.

X. Examples

A. Identification of mouse MRP

To identify novel genes acting downstream of estrogen, expressed sequence tags (EST) that represent messenger RNA (mRNA) molecules that are more abundant in the liver of normal male mice treated with estrogen than in the liver of estrogen receptor alpha knockout (ERKO) male mice also treated with estrogen were obtained by subtraction cloning. The DNA sequence of complementary DNA (cDNA) clones representing the full-length mRNA molecules corresponding to one of the above mentioned EST's was obtained and is the subject of this invention disclosure. This DNA sequence was named mMRP, for mouse myosin related protein.

Two cDNA sequences were obtained. The first, designated variant 1 is comprised of 6293 nucleotides (SEQ ID NO:2). The second, designated variant 2 is comprised of 4375 nucleotides (SEQ ID NO:3). The DNA sequence of variant 2 is identical to the first 4352 nucleotides of variant 1 indicating that variant 2 results from the use of an alternative polyadenylation site.

Translation of both variant 1 and variant 2 DNA sequences in all three frames identified a single open reading frame present in both variants that encodes a putative protein of 1337 amino acids. This putative protein starts with a methionine codon at nucleotide 125 of the cDNA sequence and terminates with a TGA stop codon at nucleotide 4138. The predicted protein sequence is shown as SEQ ID NO:1. The DNA sequence surrounding the initiator methionine codon for this putative protein matches the

criteria defined for efficient translation initiation as defined by the Kozak consensus sequence (Kozak, M. (1989) *J. Cell Biol.* 108:229-241). For most mRNAs translation initiates at the first AUG codon at the 5' end of the mRNA. In the case of the cDNA sequences for mMRP there is a single ATG codon upstream of the ATG codon that initiates the 1337 amino acid protein. This upstream ATG also lies in a favorable sequence context for translation initiation, and is followed by an in-frame stop codon. In this case, the ribosome may initiate transcription at the first AUG, terminate at the in-frame stop codon and then continue to scan down the mRNA until it reaches the second AUG codon where translation is re-initiated. Alternatively, upstream AUG codons may lie within introns as has been shown for some genes, in which case the cDNA may represent a partially processed transcript in which this intron was not removed. This suggests that the predicted protein shown as SEQ ID NO:1 is the protein encoded by this cloned cDNA.

Searching the public DNA and protein sequence databases (GenEmbl and Swissprot) with both the DNA sequences (SEQ ID NOs.:2 and 3) and the putative encoded protein (SEQ ID NO:1) reported here, failed to find an exact match. A low level of sequence similarity to the partial coding sequence for mouse unconventional myosin MY015 (accession number AF053130) was found at the protein level (31% identity over 747 amino acids). The cDNAs shown here as SEQ ID NOs: 2 and 3, and the encoded protein (SEQ ID NO:1), were therefore designated as mouse un-conventional myosin related protein (mMRP).

A search of the mMRP protein using hidden Markov models (HMM) identified several regions with similarity to known protein domains. Amino acids 346 to 458 of mMRP are similar to the MyTH4 domain which was identified as a highly conserved domain within the tail regions of four divergent members of the myosin superfamily (Chen *et al* (1996) *Genomics* 36:440-448). Two sequences with homology to myosin family IQ motifs were detected at amino acid positions 1610171, and at 180-190. A sequence with similarity to Src homology domain 3 (SH3) was identified within amino acids 1247 to 1303 of mMRP. SH3 domains have been implicated in protein-protein interactions and are present in a number of signaling molecules, indicating that mMRP may function as a novel signaling protein within cells. SH3 domains have been identified in several of the class I un-conventional myosins , and this, together with the presence of the Myth4 domain, supports the idea that mMRP is a divergent member of the myosin superfamily. All members of the myosin superfamily posses a moderately conserved

region of some 800 to 1000 amino acids at the amino terminus called the myosin head. Within the myosin head, binding sites for ATP and actin have been identified, and these sites are highly conserved among different family members. Although the amino terminal 700 amino acids of mMRP exhibit significant homology to the head domains of un-conventional myosins, the ATP-binding and actin-binding domains are absent, suggesting that mMRP may function differently from previously identified myosins.

Northern blot analysis with RNA prepared from the livers of normal male mice treated with estrogen and ERKO male mice treated with estrogen and a hybridization probe derived from the mMRP cDNA demonstrated that expression of mMRP mRNA requires the presence of a functional estrogen receptor α, and indicates that transcription of mMRP in the liver is activated by estrogen specifically via the α but not the β receptor.

Analysis of the expression of the mMRP mRNA in other tissues from the mouse was performed by reverse transcriptase-polymerase chain reaction (RT-PCR). These results showed that mMRP is expressed in brain, kidney, lung, liver and other tissues. The expression is generally up-regulated by estrogen, however, may through different receptors. For example, in brain, mMRP is up-regulated in ERKO mouse by estrogen suggesting ERβ mediated regulation.

B. Identification of human MRP

The DNA sequence of mouse MRP gene was used to pull out an EST from the Incyte database. Primers were designed based on the EST and were used for 5' and 3' RACE. RACE was performed on Clontech's Marathon Ready Human Fetal Kidney cDNA library using a gene specific primer of 5' or 3' orientation and AP1 primer supplied by Clontech. A touchdown method of PCR was used in a Perkin Elmer 9600 machine using 72°, 70°, and 68°C as annealing temperatures along with Advantage PCR Mix (Clontech) for a hot start. RACE results showed multiple bands which then were isolated and gel purified using GeneClean (Bio101). The fragments were then Taq polished and shotgun cloned into TOPO PCR 2.1 (Invitrogen). Two 3' RACE clones were identified by restriction digest and sequencing. They differ only in the middle part of the sequences by a gap of 394 base pairs. Both nucleotide sequences then merge again and terminate in poly-A tails, suggesting that they are splicing variants of each other. The 5' RACE clone was identified by sequence overlapping with the 3' clones.

Amino acid homology between the partial human and full-length mouse MRP sequences is 61%. The homology over the region corresponding to amino acids 915-1319 o f the mouse sequence (SEQ ID NO:1) is 74.5%. A comparison of amino acids 914-1320 of the mMRP amino acid sequence with amino acids 1-405 of the human MRP sequences is shown in Figure 8.

It is understood that the embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.